European Commission | Horizon 2020 European Union funding for Research & Innovation

Big Data technologies and extreme-scale analytics

Multimodal Extreme Scale Data Analytics for Smart Cities Environments

D6.2: Evaluation Report[†]

Abstract: This document provides a detailed report on the evaluation activities carried out up to the first integration phase (R1) of the MARVEL framework. The evaluation is carried out against a set of pre-defined parameters and KPIs, that had been organised in a number of categories during the earlier stages of the project. This document reports on (a) Project-related KPIs, (b) Use Case specific KPIs, and (c) Asset-specific KPIs. The document tabulates and analyses the results and evaluations of various components and goals and provides suggestions for the R2 integration phase. In addition, this document (D6.2) forms the basis for the final version (D6.4) that will also include evaluation and analysis of the framework KPIs, business KPIs and societal goals that consider the impact of the project as a whole.

Contractual Date of Delivery	31/10/2022
Actual Date of Delivery	06/11/2022
Deliverable Security Class	Public
Editor	*Adrian Muscat (GRN)*
Contributors	All *MARVEL* partners
Quality Assurance	*Thomas Festi (MT)* *Alessio Brutti (FBK)*

[†] *The research leading to these results has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957337.*

The *MARVEL* Consortium

Part. No.	Participant organisation name	Participant Short Name	Role	Country
1	FOUNDATION FOR RESEARCH AND TECHNOLOGY HELLAS	FORTH	Coordinator	EL
2	INFINEON TECHNOLOGIES AG	IFAG	Principal Contractor	DE
3	AARHUS UNIVERSITET	AU	Principal Contractor	DK
4	ATOS SPAIN SA	ATOS	Principal Contractor	ES
5	CONSIGLIO NAZIONALE DELLE RICERCHE	CNR	Principal Contractor	IT
6	INTRASOFT INTERNATIONAL S.A.	INTRA	Principal Contractor	LU
7	FONDAZIONE BRUNO KESSLER	FBK	Principal Contractor	IT
8	AUDEERING GMBH	AUD	Principal Contractor	DE
9	TAMPERE UNIVERSITY	TAU	Principal Contractor	FI
10	PRIVANOVA SAS	PN	Principal Contractor	FR
11	SPHYNX TECHNOLOGY SOLUTIONS AG	STS	Principal Contractor	CH
12	COMUNE DI TRENTO	MT	Principal Contractor	IT
13	UNIVERZITET U NOVOM SADU FAKULTET TEHNICKIH NAUKA	UNS	Principal Contractor	RS
14	INFORMATION TECHNOLOGY FOR MARKET LEADERSHIP	ITML	Principal Contractor	EL
15	GREENROADS LIMITED	GRN	Principal Contractor	MT
16	ZELUS IKE	ZELUS	Principal Contractor	EL
17	INSTYTUT CHEMII BIOORGANICZNEJ POLSKIEJ AKADEMII NAUK	PSNC	Principal Contractor	PL

Document Revisions & Quality Assurance

Internal Reviewers

1. *Thomas Festi, (MT)*
2. *Alessio Brutti, (FBK)*

Revisions

Version	Date	By	Overview
0.2.3	5/11/2022	Adrian Muscat	Updates to further PC comments
0.2.2	3/11/2022	Adrian Muscat	Updates following PC comments
0.2.1	29/10/2022	Adrian Muscat	Updates following second review
0.2.0	28/10/2022	Adrian Muscat	Updates following review
0.1.1	18/10/2022	Adrian Muscat	Document to Reviewers
0.1.0	17/10/2022	Adrian Muscat	Collated 1st Draft Document
0.0.3	03/10/2022	Adrian Muscat	Updated ToC
0.0.2	22/09/2022	Adrian Muscat	Updated ToC
0.0.1	12/09/2022	Adrian Muscat	Updated ToC
0.0.0	08/09/2022	STPM, WPL	Comments on the ToC.
0.0.0	18/08/2022	Adrian Muscat	ToC Draft

Disclaimer

The work described in this document has been conducted within the MARVEL project. This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957337. This document does not reflect the opinion of the European Union, and the European Union is not responsible for any use that might be made of the information contained therein.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

AAC	Automated Audio Captioning
AB	Advisory Board
AI	Artificial Intelligence
API	Application Programming Interface
ASC	Acoustic Scene Classification
AudioAnony	Audio Anonymisation
AV	Audio-Visual
AVAD	Audio-Visual Anomaly Detection
AVCC	Audio-Visual Crowd Counting
AVDrone	Drone-based Audio-Visual data collection
AT	Audio tagging
AUC	Area Under the ROC Curve
CATFlow	Data Acquisition Framework
CCTV	Closed-Circuit Television
COVID-19	Coronavirus Disease – 2019
CPU	Central Processing Unit
D#.#	Deliverable #.#
DatAna	Data Acquisition Framework
DFB	Data Fusion Bus
DISCO	auDiovISual Crowd cOunting dataset
DL	Deep Learning
DMT	Decision Making Toolkit
DPO	Data Protection Officer
DynHP	Compressed Models
E2F2C	Edge to Fog to Cloud
EAB	External Advisory Board
EB	Ethics Board
EdgeSec	Security Services at the edge
ELAN	EUDICO Linguistic Annotator
FedL	Framework and implementation of ML algorithms – Federated learning
FL	Federated Learning
FLOPS	Floating Point Operations Per Seconds
FPS	Frames Per Second

GA	Grant Agreement
GAN	Generative Adversarial Network
GB	Gigabyte
GDPR	General Data Protection Regulation
GPS	Global Positioning System
GPU	Graphics Processing Unit
GPURegex	GPU Pattern Matching Framework
GUI	Graphical User Interface
H2020	Horizon 2020 Programme
HDD	Hierarchical Data Distribution
HPC	High Performance Computing
HTTP	HyperText Transfer Protocol
ICT	Information and Communication Technology
IoT	Internet of Things
IP	Internet Protocol
IT	Information Technology
KPI	Key Performance Indicator
M#	Month #
mAP	mean Average Precision
MAE	Mean Absolute Error
MB	Megabyte
MEMS	Micro Electro-Mechanical Systems
ML	Machine Learning
MQTT	Message Queuing Telemetry Transport
MVP	Minimum Viable Product
NUS	Non-Uniform Sampling
O#	Objective #
OS	Operating System
PC	Project Coordinator
R#	Release
R	Report
RAM	Random Access Memory
REST	REpresentational State Transfer
RPi	Raspberry Pi
RTSP	Real-time Streaming Protocol

S2S	Site-to-Site
S3	Simple Storage Service
sec	second
SED	Sound Event Detection
SED@Edge	Sound Event Detection at the Edge
SELD	Sound Event Localisation and Detection
SET	Sound Event Tagging
SmartViz	Advanced Visualisation Toolkit
SOTA	State-of-the-Art
T#.#	Task #.#
TAD	Text Anomaly Detection
ToC	Table of Contents
UC#	Use Case
UCSD	User Centered System Design
UI	User Interface
USB	Universal Serial Bus
VAD	Voice Activity Detection
VCC	Visual Crowd Counting
ViAD	Visual Anomaly Detection
VideoAnony	Video Anonymisation
VM	Virtual Machine
VPN	Virtual Private Network
WiFi	Wireless Fidelity
WP#	Work Package #
Y#	Year #

Executive Summary

Smart city environments generate large amounts of data from multimodal sources such as video cameras and microphones installed across the cities. Most of this data is largely underutilised and eventually deleted, mainly because of engineering and technology limitations. In an attempt to narrow the gap, project MARVEL is developing an experimental framework to manage the flow and processing of multimodal data over an Edge-to-Fog-to-Cloud (E2F2C) infrastructure, which would allow the end-user (e.g., researchers, engineers, managers or policy-makers) to extract useful information from the raw data via a graphical user interface (GUI). The platform will be demonstrated in three trial cases implemented in the Municipality of Trento (Italy), the Island of Malta (Malta), and the City of Novi Sad (Serbia).

Within the project MARVEL, experiments, verification, and validation is an ongoing process and forms part of an iterative process that drives the progress of the project. This document reports the evaluation results up to the first prototype release (R1), which concerns the implementation of five use cases across the island of Malta, the municipality of Trento and the city of Novi Sad.

The document first summarises the aim and objectives of the MARVEL project and provides a summary of the evaluation methodology, which was projected during the earlier stages of the project (D1.2). The evaluation is carried out against a set of pre-defined parameters and KPIs, organised in a number of categories; (a) Project-related KPIs, (b) Use Case specific KPIs, (c) Asset-specific KPIs, (d) MARVEL framework KPIs, (e) Business KPIs and (f) societal goals that consider the impact of the project as a whole.

This document is written from the use case point of view and is organised mainly in separate sections per use case, whilst evaluations and parameters that cut across use cases are also included. As such, the evaluations reported in this document mostly belong to the first three categories, i.e., (a) Project-related KPIs, (b) Use case specific KPIs, and (c) Asset-specific KPIs which can be considered as unit evaluations. Nonetheless, some evaluations or groundwork related to the latter three KPIs (MARVEL framework KPIs, Business KPIs, and societal goals) which consider the wholesome functionality and impact of the project, have been carried out providing feedback to the iterative design process. This is important since the impact of the project depends on how individual components function together.

The document tabulates and discusses the results and evaluations of the various components and goals and in addition brings in feedback generated by the external evaluators (Advisory Board, Ethics Board, and Info Day (Trento)) which also had a say in shaping the R1 period. The experience gathered in this document will be used as feedback to the R2 integration phase, during which five more use cases will be implemented and the framework-wide KPIs can be evaluated.

1 Introduction

Project MARVEL considers smart cities as multimodal data pools, where data is generated from a very large number of IoT devices and sensors, such as video cameras and microphones. One of the biggest engineering challenges is the extraction of useful knowledge from this data. This presents an opportunity for developing new methodologies, techniques, and tools for information extraction and manipulation that differ from the traditional ones. MARVEL, therefore, aims to develop and harmonise techniques and technologies in the areas of Artificial Intelligence, multimodal perception, software engineering, High Performance Computing, and system architectures to process heterogeneous and distributed data in smart city environments. The project is therefore developing an experimental prototype, the MARVEL framework, that manages the flow and processing of data over a complex Edge-to-Fog-to-Cloud (E2F2C) infrastructure. The platform receives data from a variety of IoT devices and allows the consumer (researchers, engineers, managers or policy-makers) to process the data via a Graphical User Interface (GUI). The platform will be showcased and evaluated over several use cases executed in the municipality of Trento (Italy), the island of Malta, and the city of Novi Sad (Serbia).

Work package WP6 is responsible to oversee the execution of the use cases and the evaluation of each, mainly from a user point of view and in so much as how the societal challenges are addressed. To do so it is also necessary to follow the progress in the technical components and the framework as a whole. The evaluation falls under the remit of task T6.3 and this document (D6.2) reports the findings.

1.1 Purpose and scope of this document

The evaluation is carried out against the methodology defined in D1.2[1]. In summary, all KPIs are monitored for each experiment in both operational and technical terms, by utilising processes that allow the systematic recording of the necessary information. This information is captured under standard operating conditions and is mainly provided by data providers and experiment owners, after consultation with the assets and framework partners. In addition, an impact analysis is carried out for each experiment and the feedback from the external evaluators is taken into consideration. The information gathered, and the analysis carried out are reported in this document.

More specifically the objectives of this document are to:

1. Evaluate the use cases from a user point of view
2. Evaluate to what degree societal challenges are addressed
3. Summarise the evaluation of the technological assets and whether they are meeting the use case requirements
4. Summarise the overall project-related KPIs
5. Provide suggestions for the next phase of implementation and experimentation.

1.2 Intended Readership

This document is intended for a variety of readership groups, including potential end-users interested in the various use cases, implemented (sections 3 - 7), smart city stakeholders

[1] MARVEL D1.2: MARVEL's experimental protocol, 2021. Confidential.

interested in upcoming and cutting-edge technology (sections 2-7), and researchers from academia or industry developing or investigating similar solutions (Section 8).

1.3 Contribution to WP6 and project objectives

WP6 is concerned with ensuring the implementation and evaluation of the real-life societal experiments in smart city environments, or use cases, implemented in the municipality of Trento (Italy), the island of Malta, and the city of Novi Sad (Serbia).

The WP is organised across three tasks. The first task, T6.1, aligns the project activities with the experimental protocol (first developed in WP1, it includes execution timeframes, evaluation scenarios and selection of tools) to ensure the efficient execution of the use cases (i.e., the real-life societal experiments in smart city environments). Furthermore, it follows any adaption required in the execution of the use cases. The second task, T6.2, oversees the implementation of the use cases and ensures that all the steps in the data value chain are implemented for each use case. D6.1[2] defines the implementation of each use case separately and for each use case tabulates (a) all components or assets that are used in the implementation, (b) the specific datasets relevant to the use case, (c) user stories, and (d) the overall evaluation use case parameters for the GRN3: Traffic Conditions and Anomalous Events, GRN4: Junction Traffic Trajectory Collection, MT1: Monitoring of Crowded Areas, MT3: Monitoring of Parking Places, UNS1: Drone Experiment, use cases. In addition, use case specific functional and non-functional KPIs and asset KPIs (D1.2) are revised in sections 2.6.5 and 2.6.6 respectively.

Finally, T6.3 is concerned with the evaluation and impact analysis of the use cases, which also requires input from the technical evaluation of assets and framework, which is normally carried out in WP3 and WP5. This document (D6.2) gathers the information related to the evaluation of the parameters and KPIs in the respective tables (defined in D1.2 and refined in D6.1), carries out a cross-table analysis and provides suggestions for the second period of integration (R2). This follows the development process in MARVEL, which is largely iterative, i.e., early versions of the framework may have limited functionality and more functionality is added as required in later versions. Similarly, the performance and adaptation of the individual components may take place in later versions of the framework.

1.4 Relation to other WPs and deliverables

This deliverable is directly related to D1.2 (“MARVEL’s Experimental protocol”) and D5.2 (Technical evaluation and progress against benchmarks – initial version)[3].

D1.2 describes the initial evaluation and benchmarking strategies (section 6.3.1) and classifies and provides definitions for all types of KPIs (sections 6.3.3, 6.3.4 and 6.3.5), which address the use case, asset, and MARVEL framework KPIs respectively. In addition, D1.2 provides an initial attempt to benchmark business metrics (Table 7.1) and societal goals and benchmarks (Table 7.2).

D5.2 (the initial version) and eventually D5.5 (the final version) document the implementation of the benchmarking strategies. The results in these documents are reflected in this deliverable (D6.2). The final version (D5.5) document, which will contain the benchmarking of the full MARVEL framework, will be delivered at M30 on time to be used for D6.4 (the final version

[2] MARVEL D6.1: Demonstrators execution - initial version, 2022. https://doi.org/10.5281/zenodo.6862995

[3] MARVEL D5.2: Technical evaluation and progress against benchmarks – initial version, 2022. https://doi.org/10.5281/zenodo.6322699

of D6.2), which will be delivered at M36. D5.4[4] also refers to the demonstration of R1 and presents a preparatory roadmap for the MARVEL benchmarking activities.

Tasks T7.3 and T7.5 in WP7 deal with the projection of exploitation activities, the long-term sustainability and IPR management for potential commercialisation. These activities will evaluate business opportunities for project MARVEL.

1.5 Structure of the document

This chapter outlined the scope of this document and how it is related to other deliverables and WPs in the project. Chapter 2 gives a summary of the MARVEL project, a list of use cases, societal challenges and the overall MARVEL objectives, an overview of the evaluation and benchmarking strategy and the types of KPIs and evaluation parameters. Chapters 3-7 describe the five use cases implemented and the respective evaluation information that has been gathered and that will be obtained during the R2 period. Chapters 8-9 discuss the asset-specific KPIs, the project-wide KPIs, the business KPIs and the MARVEL framework KPIs. Chapter 10 describes in detail the input from the Advisory Board (AB), the Ethics Board (EB), and the external stakeholders. Chapter 11 summarises and concludes the report.

[4] MARVEL D5.4: MARVEL Integrated framework – initial version, 2022.

2 The MARVEL project and its Evaluation

This chapter provides the reader with a summary of background information on the MARVEL project prior to delving into the evaluation chapters of the document.

2.1 Summary of the MARVEL Project and Framework

This subsection is a summary of what MARVEL aims to achieve. The purpose is to provide a background on the main objectives of MARVEL, the technical framework and its components, and the real-world demonstrators. For a more detailed treatise, the reader is referred to documents D1.3[5] and D6.1.

The overall MARVEL objective is to research and develop a framework that can efficiently process data collected from smart city IoT devices and sensors and produce information that is useful in addressing societal challenges. The framework is demonstrated over several use cases covering road traffic, city open spaces and large events monitoring and management to address societal challenges in sustainable mobility and climate change, and security in cities. For more details on the use cases and the societal challenges addressed, the reader is referred to sections 3.2, 4.2, 5.2, 6.2, and 7.2 of this document.

To achieve its overall goal, the project is based on four main pillars; (1) *Real, heterogenous, distributed Big Data in smart cities*, i.e., the real-world use case implementation, (2) *AI-based processing for multimodal (mainly video and audio) perception and situational awareness*, (3) *an E2F2C ubiquitous computing architecture* that allows for the distributed collection and management of data, AI model training and inference, a user interface and the optimisation of resources for data processing, and (4) *Validation and quantitative assessment of the E2F2C and multimodal tools and methods via societal, business and industry-validated benchmarks*.

To realise the framework, it is necessary to build or modify and improve technological components and assets at every level of the E2F2C architecture. Figure 1 depicts an overview of the MARVEL conceptual architecture, which consists of no less than thirty-five components grouped into seven MARVEL subsystems;

1. The **sensing and perception** subsystem includes the various sensors, mainly cameras and microphones, data capturing systems, such as the AVDrone, data capturing apps (e.g., sensMiner and the asset registry).
2. The **security, privacy and data protection** subsystem is responsible for preserving privacy via the anonymisation of the audio and video sources, the implementation of end-to-end communication and device security.
3. The components of the **data management and distribution** subsystem manage and optimise the data flow across the E2F2C infrastructure.
4. The **audio, visual and multimodal AI** subsystem is responsible for providing the AI models and algorithms (for example video anomaly detection and sound event detection), with the related training processes, that are necessary to implement the use cases.

[5] MARVEL D1.3: Architecture definition for MARVEL framework, 2021. https://doi.org/10.5281/zenodo.5463897

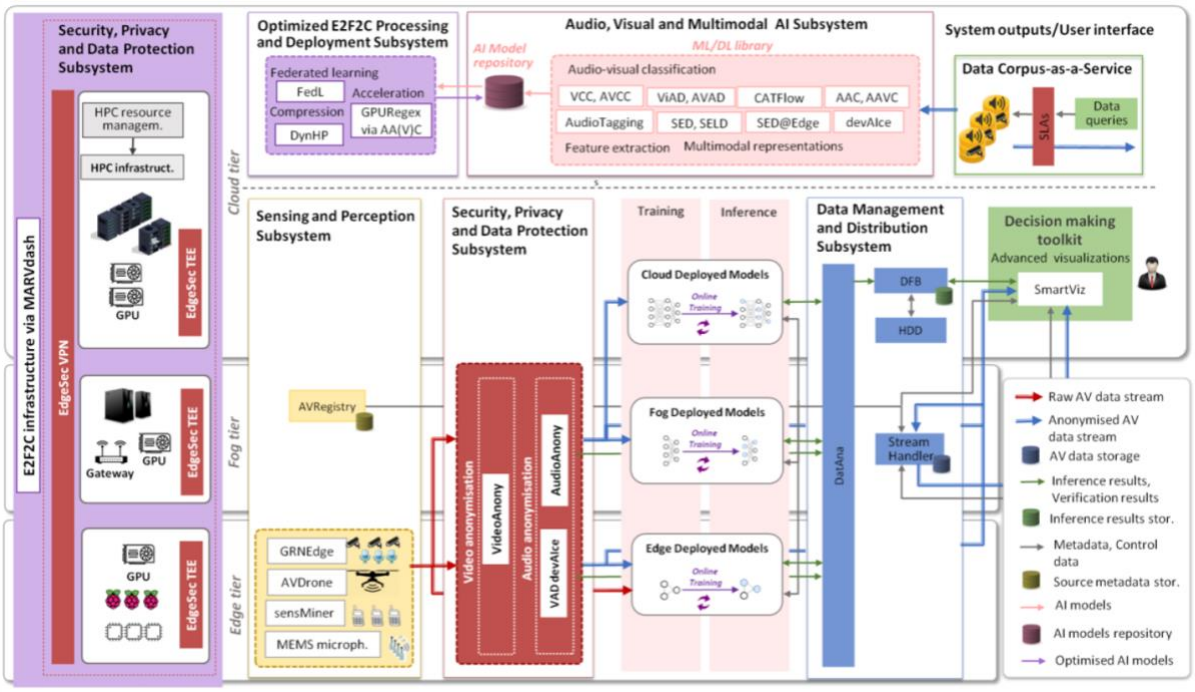

Figure 1: The MARVEL Conceptual Architecture

5. The **optimised E2F2C processing and deployment** subsystem is responsible for the delivery of optimised AI models and includes model compression, federated learning, and GPU acceleration.
6. The **E2F2C infrastructure** subsystem is the collection of inter-connected computational resources made available by the project partners and includes Intel NUC, RPis, PCs, mobile phones at the edge, workstations and local servers at the fog and HPC infrastructure at the Cloud layers.
7. The **system outputs** consists of the **Decision-making Toolkit (DMT) and the MARVEL Data Corpus** elements. The DMT is the main interface of the MARVEL framework to the end-users (traffic managers, police and law-enforcement personnel, researchers, academics and engineers) and includes a graphical user interface (GUI) that facilitates the use of the system. The MARVEL Data Corpus provides access to labelled, anonymised training data from MARVEL pilots. Figure 2 depicts an example from the GUI for the use case GRN3.

2.2 Types of KPIs and evaluation parameters

Due to the complex nature of the MARVEL framework, the KPIs have been organised in four categories; (a) Project-related KPIs, (b) Use Case specific KPIs, (c) Asset-specific KPIs, and (d) MARVEL Framework KPIs. Here we briefly summarise what the KPIs are about and how they have been defined. Further details are available in D1.2 and D6.1.

Besides the project-related KPIs reported in the GA related to the technological and scientific objectives of the project as a whole, crucial for an effective assessment of the project's success are the *use case specific KPIs*. These KPIs capture the requirements of the end-users (the pilots) and are expected to evaluate whether the project solved or not the technical and societal challenges and issues that are behind the participation of the pilots in the project. For each use case, we collected technological goals as well as societal and business expectations. They have

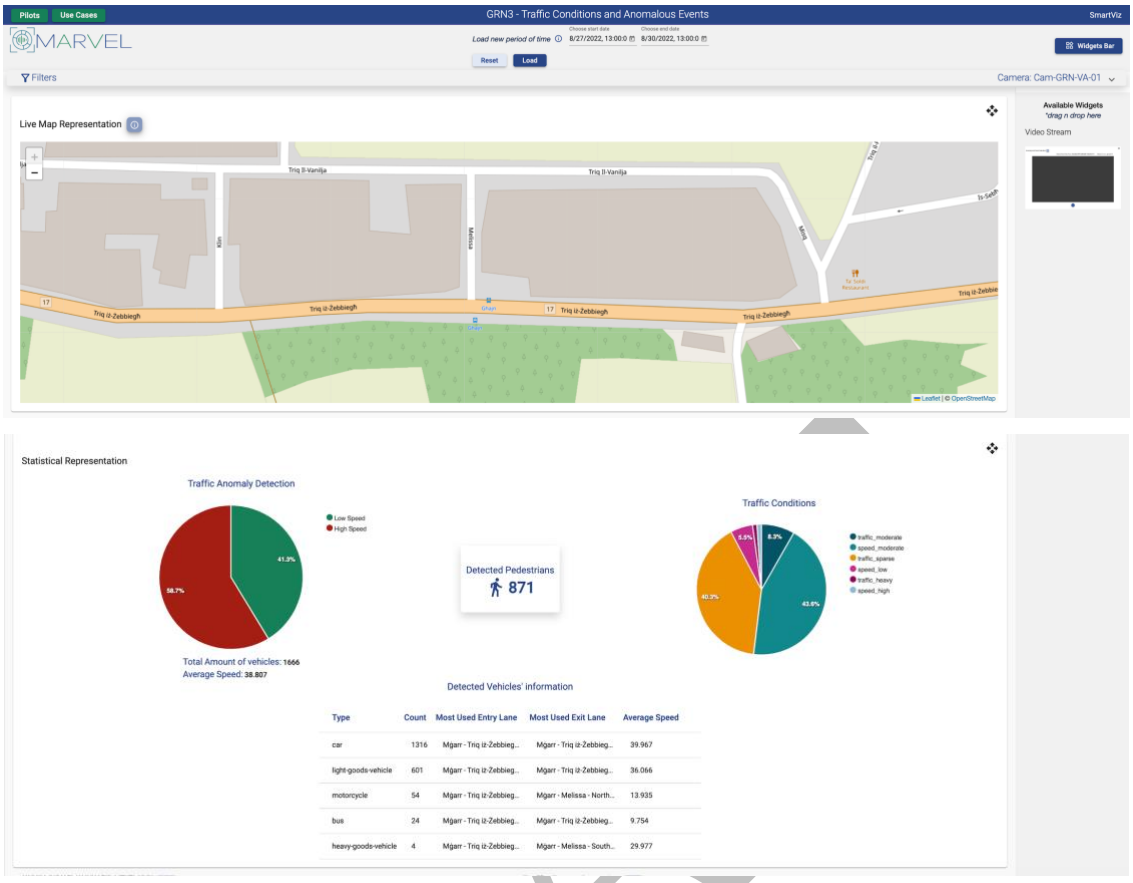

Figure 2: GUI implemented on SmartViz for the use case GRN3: Traffic conditions and anomalous events

been defined via questionnaires and have been categorised into 2 groups: *Functional and non-Functional KPIs.*

Functional KPIs measure if and how the MARVEL platform solves the pilot issues. Examples are:

- GRN-KPI1: to increase the cyclists detection rate by 10% with respect to current technology.
- MT-KP6: to reduce the reaction time in case of issues with respect to a single person monitoring all the cameras.

These objectives are in general evaluated by internal personnel of the pilots.

Non-Functional KPIs refer instead to other features of the solution, not directly related to the final technological goals. Examples are:

- usability and satisfaction for the end-users;
- acceptance and satisfaction by the population;
- scalability, costs, economic benefits.

These KPIs are particularly challenging because they involve often: external evaluators whose engagement is not always straightforward; the evaluation or estimation of long-term impacts.

In spite of the diverse nature of the pilot application scenarios, several non-functional requirements are similar across pilots. Among these cross-cutting KPIs, the GDPR-related requirements are implicitly defined for all pilots.

Besides delivering end-to-end solutions for the pilot end-user, one of the goals of the project is also to advance all the technological assets provided by the partners. These improvements were initially defined in the GA and finalised in D1.2 providing, for each asset:

- one or more target improvements;
- evaluation metrics;
- benchmarking datasets;
- comparison baselines;
- the related project KPIs.

Asset KPIs have been collected directly from each involved partner, following a bottom-up strategy. As the state-of-the-art rapidly evolves in most scientific fields involved in MARVEL, a periodic revision of these KPIs is carried out as part of the benchmarking process in WP5 and WP6.

All asset and use case KPIs have then been merged identifying the evaluation parameters and the evaluation variables of the overall *MARVEL framework*. Overall, 5 evaluation parameters were defined: performance, usability, scalability, trustworthiness and reliability; and 15 evaluation variables. For each evaluation variable, metrics and affected MARVEL subsystems were defined.

2.3 Benchmarking and Evaluation Strategy

In the context of D1.2, three types of benchmarks to assess and evaluate the MARVEL system were highlighted:

- Technical benchmarks: Assessing how the MARVEL framework and the individual components go beyond the state-of-the-art on technical excellence. This type of benchmarks is mostly related to the work carried out in the scope of WP5.
- User experience benchmarks: Looking at the usability of the system by the end-users and therefore related mostly to the "User interaction and decision-making" layer of the MARVEL architecture.
- Business (model) benchmarks: Validating the business-related aspects, such as processes, sustainability of the solutions and cost perspective. This has to do with the work carried out in exploitation (WP7) but encompasses the assessment of some business metrics that can help to understand the industrial potential of the results.

The first two will be used to assess the technical and usability aspects, while the final will help to understand the business value of the MARVEL solution.

As WP5 is going to perform dedicated technical benchmarking for the components and the MARVEL framework as a whole, the strategy in WP6 related to the technical evaluation will focus mainly on the overall perception of the system behaviour from the perspective of the fulfilment of the use cases, looking at aspects such as general performance, scalability, robustness, accountability, or privacy awareness, leaving more technical details of the various components to WP5 technical assessment.

As hinted in D1.2, usability will be assessed via dedicated questionnaires or interviews with the users involved in the use cases to validate mainly the current version of the user interface in a qualitative fashion. From the business perspective, D1.2 hinted a set of typical metrics to measure business performance. Some of these metrics are assessed to some extent in this deliverable, especially those related to efficiency (Time Efficiency, Service Quality) that are directly related to costs (Cost Reduction, Profit and Revenues Increase). However, a more in-depth assessment of the business metrics will be carried out over the final release version of the platform in collaboration with WP7 in relation to the business value and innovation of the MARVEL framework with regard to the pilots' use cases.

The sections related to the specific use cases in this document will present a summary of some of these metrics, such as Efficiency, Operability, Usability, Robustness, Performance, etc., along with an overview of specific KPIs related to the pilots. This will serve as an initial assessment of the system with regard to the specific pilot use cases developed so far, and as a basis and feedback for the final assessment to be carried out at the end of the final iteration.

3 Use case GRN3: Traffic Conditions and Anomalous Events

3.1 Introduction

This chapter gathers all the information pertaining to use case GRN3: Traffic Conditions and Anomalous Event, including evaluation progress. The first two sections summarise the scope and description of the use case (section 3.2), the components used in the use case, including relevant datasets and the configuration of the framework (section 3.3). These sections provide a background to the evaluations (relevant to the use case) reported in the sections that follow (i.e., use case specific parameters, KPIs).

3.2 Scope and description of use case

The intended application for this use case is the monitoring of traffic conditions and the detection of anomalous events, for example: traffic jams, accidents, cars stuck and obstructing a junction, very slow vehicles and service vehicles parked on the side or obstructing a carriageway. The latter event is frequent in Malta's narrow one-way urban streets, often causing ripple effects that extend beyond the immediate area. In general, this output would find application in, for example, systems intended to inform drivers near the detected anomaly or to infer possible issues in adjacent areas and inform drivers of obstacles ahead. In addition, the detection of anomalous events can be used to alert personnel stationed at traffic management control rooms, who can then manually process the data and take any necessary action. This latter application could potentially improve the efficiency of operations.

The accurate detection and detection time are two crucial metrics to evaluate this use case. The R1 integration is designed such that both metrics can be obtained. Whenever an anomaly occurs, the anomaly is flagged in the control room such that actions can be taken. In particular, a video clip of the anomaly is saved and is made available to the traffic control personnel, thus gaining visual insight into the anomaly, which can help in determining the right course of the action. This feature is essential for the evaluation of both metrics since the traffic personnel would be able to validate both the anomaly detected as well as the traffic conditions. In addition, the time stamps on the video stream can be used in the calculation of the overall system latency.

The users of this system are intended to be traffic managers who can give directives to other authorities to react to any disruptive event, for example, a traffic incident. The system would partially replace manual systems, where personnel would typically observe multiple video screens to pick up anomalies. This use case could potentially reduce the time taken to respond to disruptive events and therefore ease traffic congestion, which is one of the societal challenges.

3.3 Components, assets, and framework configuration

Figure 3 depicts the framework configuration for the GRN3 use case. Three cameras are deployed in the implementation of this use case, one at Mgarr and two at the Zejtun location. The three cameras deliver continuous AV streams with both modalities present (audio, video). The AV stream from the Mgarr camera is first transmitted to the GRN edge PC (GRN E1) where it is consumed by the VideoAnony component. In parallel, i.e., concurrently in time, the two cameras in Zejtun produce similar AV streams, both of which are transmitted to the GRN fog server (GRN F1), where two different instances of VideoAnony receive the respective streams and anonymise them, frame-by-frame. Each of the three VideoAnony instances produces an RTSP stream that is consumed by several components further up in the pipeline as follows; 1) the AI components (AVAD and AT), to produce inference results; 2) StreamHandler, for AV data storage; and, finally, 3) SmartViz, for real-time visualisation of

the AV streams. StreamHandler and an instance of AV Registry are both deployed at the fog layer. DFB, DataCorpus and Smartviz are all implemented at the Cloud layer, (PSNC HPC via OpenStack).

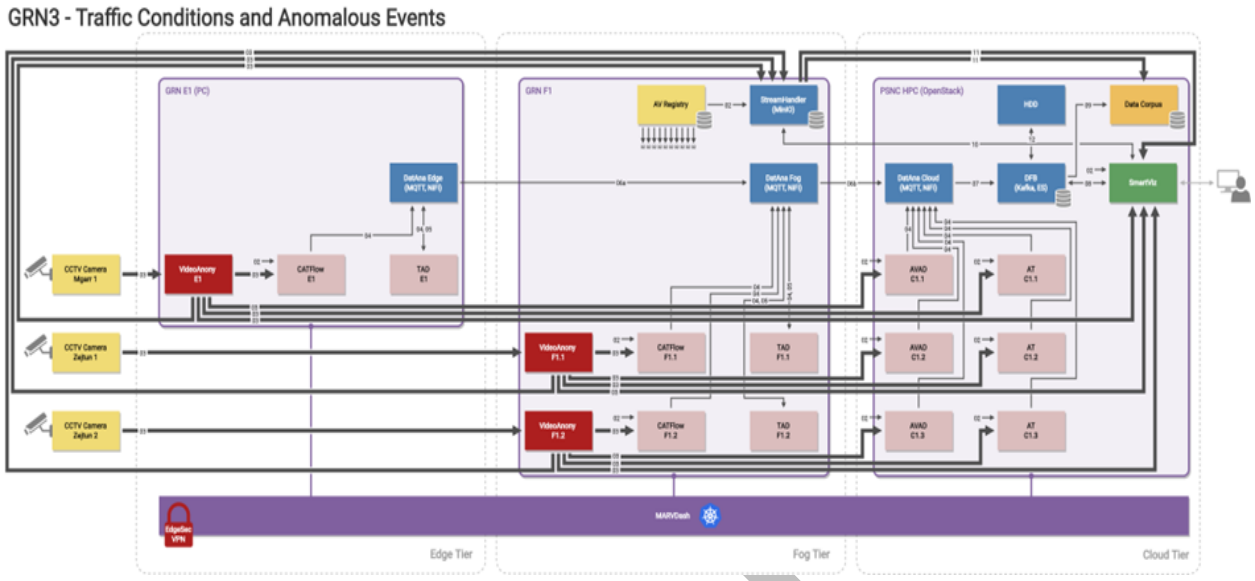

Figure 3: MARVEL R1 deployment and runtime view of the MARVEL architecture for GRN3: Traffic Anomalous Events (source: D6.1)

DatAna is deployed at all layers of the architecture including the edge. After the inference results have been generated, either at the edge, fog or cloud, they are submitted to the respective MQTT broker of the DatAna agent residing at the same node. Each of the DatAna agents, after receiving the inference results, applies the appropriate transformation to standard data models, and relays the results to the next DatAna agent in the infrastructure, in the case of DatAna edge or DatAna fog, or sends these to the appropriate Kafka topics of. More specifically:

- DatAna edge (GRN E1) receives inference results of CATFlow+TAD at edge and sends the transformed results to DatAna fog;
- DatAna fog (GRN F1) receives the results from DatAna edge and forwards them to DatAna cloud;
- DatAna fog (GRN F1) concurrently receives inference results of CATFlow+TAD at fog and sends the transformed results to DatAna cloud;
- DatAna cloud (PSNC HPC via OpenStack) receives the results from DatAna fog and sends them to the Kafka topics of DFB;
- DatAna cloud (PSNC HPC via OpenStack) receives the results from AVAD and AT instances at the cloud, transforms them, and sends results to the appropriate Kafka topics of DFB.

The AI components deployed in this use case are the following;

- **CATFlow+TAD** –The goal of the Text Anomaly Detection component is to detect anomalies in vehicle trajectories after CATFlow processes the anonymised AV stream.
- **AVAD** – The Audio-Video Anomaly Detection component is trained on normal traffic situations such that it can recognise any deviation from normal in previously unseen data, i.e., events that are not present in the training data are flagged as anomalous.
- **AT** –The Audio Tagging component is employed in GRN3 to enable tagging of audio traffic events at equally spaced time intervals.

GRN has provided data for the training of AT, AVAD, and VideoAnony, organised in three datasets. The GRN-AV-traffic-state dataset is used for training the AT model. For this dataset, AV clips from various locations and the GRN static cameras were cropped into 5–second clips and subsequently annotated with labels of specific classes, for example, low speed or moderate speed. The AVAD dataset on the other hand consisted of Audio-Video data that does not include anomalies (training set) and Audio-Video data that included anomalies (test set). Anomalies included temporary obstructions on the road, bicycles and U-turns. The training of VideoAnony required an annotated dataset with bounding boxes indicating the vehicles' number plates. The data was obtained from various locations around Malta, in addition to videos obtained from the GRN static camera.

3.4 KPI tables

This section tabulates the evaluations carried out from the point of view of the use case. These tables are sourced from D1.2 and D6.1 and the evaluations are specific to the use case.

3.4.1 Use Case Evaluation Parameters

This section summarizes (Table 1) the progress in evaluating the use case against a set of general parameters. Most of these criteria are dependent on the use case KPIs (Table 2) and the asset specific KPIs (Chapter 8).

Table 1: Parameters for use case GRN3: Traffic Conditions and Anomalous Events

Parameter	How to measure	Target to be achieved
Efficiency *Related to the efficiency of the system as used in the use case*	Efficiency is largely dependent on how long it takes to detect and flag an anomaly. The average time in minutes from the start of the anomaly to detection time is measured.	2 minutes from start of the anomalous event.
	Progress, Results and Comments: For an anomaly to be detected the video input must be processed by two AI models: VideoAnony and AVAD. The following results were acquired from the AI model's benchmarking and give indications of the expected delays for the use case: VideoAnnoy requires **5.8ms per frame**. AVAD data flow currently consists of saving 10s of video, pre-processing the data for 9s and 50ms of inference per frame, as well as approximately 10ms to send a message. Thus, AVAD results in approximately **20s of delay.** The CATFlow algorithm is able to work at a rate faster than real-time, thus no latency is observed. However, the message per vehicle is transmitted when a vehicle exits the camera FoV. The total delay depends on many other factors, such as delay in the MQTT brokers, network delays and varying workload of MARVEL hardware. Thus, further investigations are required at the framework level (R2).	
Operability *Related to the ability of the components to keep functioning together*	Record downtime for any of the components (assets) along the system pipeline as a percentage of total time.	Downtime is minimised to 10% or less.
	Progress, Results and Comments No long-term tests have been performed yet (R2).	

<table>
<tr><td rowspan="2">Usability

Related to how well the system helps the users achieve a task in a given use case</td><td>Interview external experts to measure the end-user experience.</td><td>End-users find the system easy to use</td></tr>
<tr><td colspan="2">Progress, Results and Comments

An online survey to analyse the usefulness and user experience of the tool was mailed to potential respondents. The respondents had access only to a recorded demo of the tools in use and could not try it out in a live session. Below is a summary of the responses.

All four people who took the survey have experience in the transport industry. Most had experience in using automatic traffic monitoring tools, mostly using pneumatic-tubes based counters. The pneumatic tube systems used were described as slow, easily damaged, not ideal for slow moving traffic, only give volume of traffic, having issues with batteries and thus are only temporary installations and cannot be installed on all roads.

Half of the respondents were confident that they understood the system’s capabilities while the other half were not sure. The quality of the content provided on screen, interface design of the system, features of the product and overall quality of the product were all rated positively, with most users saying they were satisfied with these aspects. Most of the users were satisfied with the readability of the characters and the screen layout. All of the users agreed that the descriptions of the widgets make sense and the prompts for inputs are clear. Most users also agreed that they would recommend such a product to be used at their workplace.

The users suggested using a satellite view map with zoom in/out and making use of nicer graphics. They also noted how this tool would mostly be useful to transport authorities, law enforcement agencies or even local councils. In addition, a tool to customise anomaly detection would help.

A majority of respondents answered that they have not seen tools similar to the live map representation, anomalous speed detection, visual and audio-based anomaly detection and statistical representations, being used in the industry. On the other hand, they did come across tools such as the live video stream tool. Overall, the respondents agreed that for each widget the data is presented in an easy-to-read manner, the data gathered and visualised seems useful in their work and they can easily understand the data input.</td></tr>
<tr><td rowspan="2">Robustness

Related to how robust are the system components during the period of operation</td><td>Sustains performance in various weather conditions (Power source, sensor, and CPU board operation).</td><td>Performance sustained in most weather conditions</td></tr>
<tr><td colspan="2">Progress, Results and Comments

Most GRN hardware has been set up such that weather conditions have minimal effect on performance. On the other hand, RPi-based systems overheat quickly.

GRN Cameras: all GRN cameras have an IP67 rating which means they are waterproof and dustproof, thus are able to sustain performance in many weather conditions. No shutdowns were experienced during the summer months due to excess heat.

GRN Edge Layer: The current edge layer is installed indoors, not exposed to any weather elements with air conditioning available if necessary.

GRN Fog Layer: The Fog layer is a high-end workstation in a server room so no disruptions due to the weather are expected.

In general, the GRN hardware provided is robust.</td></tr>
<tr><td>Performance</td><td>Correctly detect various anomalous events on the road.</td><td>70% detection of anomalous events</td></tr>
</table>

Related to how well the system performs the intended task	**Progress, Results and Comments** Current TAD performance: precision of 80% when detecting vehicles moving at anomalous speeds. AVAD performance on the Zejtun camera 1 is: True Positive Rate: 64.9 %, False Positive Rate: 13.9 %, AUC: 78% The AT component is used to recognise traffic amount and speed for a given audio segment. A three-fold cross-validation setup was used in the evaluation, and evaluation was done with the mean average precision (mAP) metric. The AT component achieves mAP metric of 0.49. Further tests will be carried out on additional anomalies.	
Accountability *Related to the system being able to explain results or decisions.*	System stores video snippet of anomaly.	Number of times system fails to store video snippet, even though anomaly is flagged.
	Progress, Results and Comments This parameter will be tested at the framework level during R2 period.	
Transparency *Related to the description of the processes or algorithms that are used to generate system output*	Decision processes are described in a document.	Document availability
	Progress, Results and Comments The processes and algorithms are explained in detail in deliverables D5.4 and D3.2[6]. The former describes the data flow and how the data is transformed as it traverses the E2F2G infrastructure, while the latter describes the AI algorithms that process the data to extract information useful for the use case.	
Privacy awareness *Related to the provision of adequate governance mechanisms that ensure privacy in the use of data.*	Inspect anonymisation in a sample of stored videos detected as anomalies.	Anonymisation of all frames within AV snippet.
	Progress, Results and Comments Benchmarking results for VideoAnony yielded the following results: the AI model is anonymising videos with a precision of 0.797, recall of 0.492, 0.614 mAP@0.5 and 0.29 mAP@0.95.	

3.4.2 Use case specific KPIs

This section tabulates the progress in evaluating the use case against a set of functional KPIs (described in Table 2 and evaluated in Table 3) and a set of non-functional parameters (Table 4).

Table 2: Evaluation scenario and relevant KPIs for the GRN3: Traffic Conditions and Anomalous Events (Table 6, D6.1)

Evaluation Scenario	Target	Relevant KPI
Testing the various AI models on a labelled dataset to determine the detection rate and F1 score achieved by the models.	The aim is to obtain a 70% detection of anomalous events.	**GRN-KPI4: Correctly detect various anomalous events on the road**

[6] MARVEL D3.2: Efficient deployment of AI-optimised ML/DL models – initial version, 2022. https://doi.org/10.5281/zenodo.6821232

Evaluation Scenario	Target	Relevant KPI
Observing the time taken to detect an anomaly through measuring the system's latency.	The aim is to detect anomalies 2 minutes after the start of the event.	**GNR-KPI5: Detection time**

Table 3: Use case specific KPIs for GRN3: Traffic Conditions and Anomalous Events (Table 6.2 in D1.2)

KPI	Metric	Baseline	Expected result/ Improvement	Evaluators
GNR-KPI4: Correctly detecting various anomalies on the road	Detection rate/F1	Labelled Dataset	70% detection of anomalous events	GRN Developers External traffic experts
	Progress, Results and Comments	Current TAD performance: precision of 80% when detecting vehicles moving at anomalous speeds. AVAD performance on the Zejtun camera 1 is: True Positive Rate: 64.9 % False Positive Rate: 13.9 % AUC: 78% The results are close to what is expected. However, further tests are required on various classes of anomalies.		
GRN-KPI5: detection time	Time in minutes	No baseline	2 minutes from start of the anomalous event	GRN Developers
	Progress, Results and Comments	Following AI model tests; AVAD results in a maximum of 20msec of delay. However, the framework latency still needs to be factored in. This parameter will be tested at the framework level during R2 period.		

Table 4: Use case specific non functional evaluation variables for GRN3: Traffic Conditions and Anomalous Events (6.3 in D1.2)

Evaluation variable	How to measure	Internal evaluators	External Evaluators
End-user experience	Survey	GRN Managers	Transport Experts (Traffic Control Room Personnel)
	Progress, Results and Comments	An online survey to analyse the usefulness and user experience of the tool was mailed to potential respondents. A summary of the results is given in Table 1, above. In general, the system was found to be interesting and useful by experts. They also provided some suggestions.	
Scalability	Cost to add new video/audio feed	GRN Managers	
	Progress, Results and Comments	**Financial Costs to set up new Camera node:** IP camera with A/V Stream Cost: 150-250 euros (if not already available) Internet Access: 25 euros/month per node Computational Cost to ingest stream with VideoAnony, CATFlow+TAD, AVAD, and AT: VideoAnony: 20GB Storage CATFlow + TAD: 50GB Storage, 1-2 CPUs, 3-4GB RAM	

		AVAD: 10GB Storage, 4 CPU cores, 16GB RAM AT:4-16 GB Storage, 1-4 CPU cores, 8GB RAM Note: all components require a GPU for best Performance.

4 Use case GRN4: Junction Traffic and Trajectory Collection

4.1 Introduction

This chapter gathers information pertaining to use case GRN4: Junction Traffic and Trajectory Collection, including progress in evaluations. The first two sections summarise the scope and description of the use case (section 4.2) and the components, including relevant datasets, and the configuration of the framework (section 4.3). These sections provide a background to the evaluations (relevant to the use case) reported in the sections that follow (i.e., use case specific parameters, KPIs).

4.2 Scope and description of use case

Junction Traffic Trajectory collection (GRN4 use case) is focused on the requirement of long-term data analytics that shed light on both the behaviour of road users (e.g., car drivers, motorcyclists, cyclists, pedestrians, etc.) and on gathering traffic statistics at road network junctions. This use case is of interest for long-term transport planning and evaluation. In particular, there is currently significant interest in studying active travel modes, such as cycling, walking, and micro-mobility, more generally. Authorities in Malta are interested in, for example, finding the optimal position of pedestrian crossings, whether provisions for cyclists at complex junctions are adequate, and whether installed provisions are being used as intended. Sustainable mobility is the societal challenge addressed by this use case.

This use case requires entity detection and its trajectory across a junction or road segment and descriptive statistics of network junction traffic. It, therefore, follows that entity detection and tracking models can be potentially used as a first processing stage, followed by further processing to generate descriptive statistics. The innovation that we are targeting, with this use, case is the construction of a queryable database that can be used to look up historical data on the trajectories of vehicles and pedestrians at junctions, with sufficient accuracy to automatically detect the query. The trajectories and data generated from the CATFlow algorithm are saved on the MARVEL Data Corpus such that the data can be accessed and processed by the end-user. Currently, the query paths can be detected through visual inspection of the trajectories, which is a feature of the system. Future integrations will have tools that automatically detect anomalous paths. This use case was also partially implemented for the Minimum Viable Product (MVP) in M12. The progress since the MVP implementation included the incorporation of all the components as part of the integrated system.

The users for this system are intended to be traffic engineers who need data to make informed decisions about infrastructure changes and upkeep, as well as transport researchers.

4.3 Components, assets, and framework configuration

Figure 4 is a diagram of the GRN4 frame configuration, which gives information on how the various components interact with each other. As can be inferred, the architecture is similar to that of GRN3, however, some of the components are different, namely the AI models. For a detailed description of the basic framework, the reader is referred to section 3.3. In addition to the basic framework, GRN4 makes use of the full CATFlow asset and the SED and AVCC AI components. As such, the DatAna cloud component receives the results from SED and AVCC instances at the cloud, transforms them, and sends results to the appropriate Kafka topics of DFB. Following is a description of the AI components deployed in GRN4.

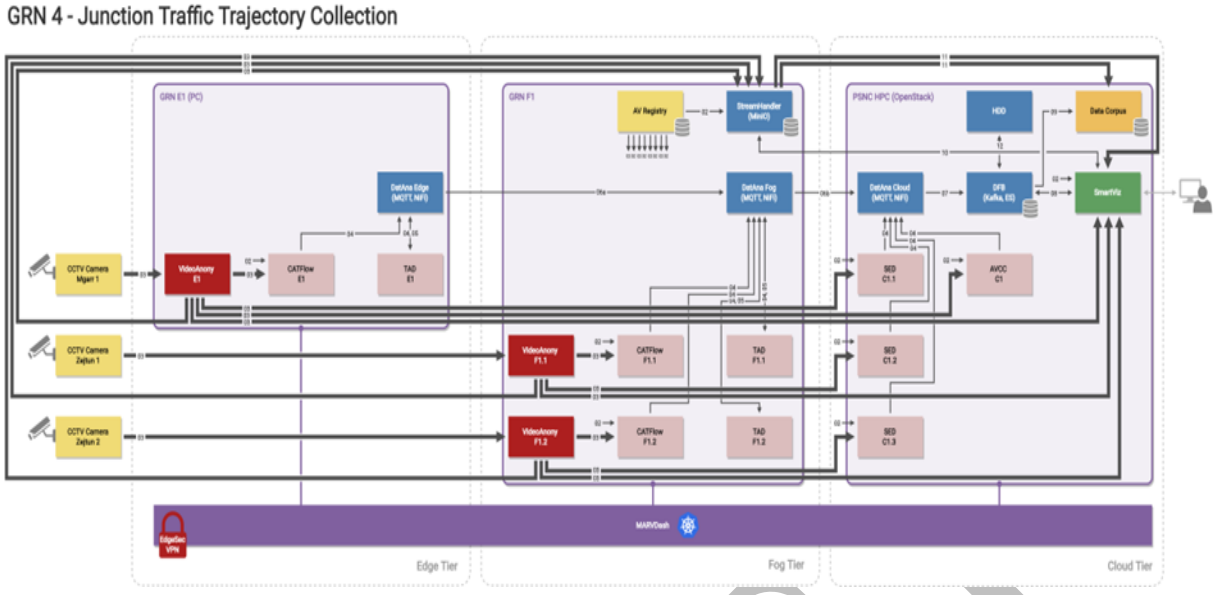

Figure 4: MARVEL R1 deployment and runtime view of the MARVEL architecture for GRN4: Junction Traffic Trajectory Collection (source: D6.1)

- **CATFlow + TAD** - The distinction with respect to GRN3 is that in GRN4 CATFlow classifies and tracks both vehicles and pedestrians. In addition, for implementation in GRN4, CATFlow is internally configured such that it outputs the trajectories of each vehicle as well as the entry and exit points within the camera field of view (FoV).
- **SED** – The Sound Event Detection component is employed in GRN4 for the detection and classification of vehicles, such classes such as cars, buses, and motorcycles, from the audio stream. Three instances of SED are deployed at the cloud (PSNC HPC via OpenStack) to process all three microphones at the Mgarr and the Zejtun locations.
- **VCC** –The Visual Crowd Counting component is deployed to process the Mgarr video stream to estimate the number of people present in the visual live feed. Crowd counting is the problem of identifying the number of people present in a visual scene. The technical details are given in D3.1[7]. The aim in the GRN4 is to count the pedestrians walking through the frame, thus determining if the area being monitored requires additional pedestrian infrastructure. The VCC can also be used to count the number of pedestrians waiting at the bus stop, thus providing data for sustainable mobility research.

GRN-AV-traffic-entity dataset is used for SED training. For this dataset audio-visual snippets were obtained from various locations around Malta and from the GRN static cameras. The audio track was manually annotated. All AV clips are 3-5 minutes long. Approximately 180 snippets were annotated for this dataset, 44 from different locations around Malta and the rest from the GRN static cameras. This amounts to more than 13 hours of annotated data. Although this amount of data exceeds the 4 hours planned in D1.2, more data will be required due to the complex requirements for training the SED model.

In addition, GRN contributed 71 snippets, that contain pedestrians, to the MARVEL AVCC dataset. The AVCC dataset is annotated with the position of pedestrians' heads. These types of annotations are very time-consuming since a short video involves a large number of frames.

[7] MARVEL D3.1: Multimodal and privacy-aware audio- visual intelligence – initial version, 2022. https://doi.org/10.5281/zenodo.6821318

The CVAT[8] software was used to streamline the process as much as possible. This software partially automates the annotation process by guessing the location of the pedestrian's head in the next frame.

4.4 KPI tables

This section tabulates the evaluations carried out from the point of view of the use case. These tables are sourced from D1.2 and D6.1 and the evaluations are specific to the use case.

4.4.1 Use Case Parameters

This section tabulates (Table 5) the progress in evaluating the use case against a set of general parameters. Most of these criteria are dependent on the use case KPIs (Table 6) and the asset specific KPIs (Chapter 8).

Table 5: Parameters for use case GRN4: Junction Traffic and Trajectory Collection

Parameter	How to measure	Target to be achieved
Efficiency *Related to the efficiency of the system as used in the use case*	Increased efficiency in the planning of roads and network infrastructure. The feedback from external traffic experts will be collected through a survey to determine the increase in efficiency	Potential decrease in time for planning through the availability of data
	Progress, Results and Comments: An online survey to analyse whether the tool improves efficiency if introduced in work practices was mailed to potential respondents. The respondents had access only to a recorded demo of the tool in use and could not try it out in a live session. Below is a summary of the responses. Two experts took the survey, one transport researcher and one in traffic data recording. In their experience, similar tools to the MARVEL framework are very difficult to acquire. One of the respondents was not confident he/she understood the capabilities of the system. They had no knowledge of any tool similar to the Temporal Representation and Detailed Temporal representation. Whilst they believe that these two tools will be useful when making decisions about traffic management, they weakly disagree that these specific tools will help them make more informed decisions in long term planning. They have seen tools similar to the Statistical Representation widget and Trajectories. Collectively, they cautiously think that these tools will be useful when taking decisions on traffic management and will help them make more informed decisions or increase the efficiency in the decision process. When used in planning, these tools are definitely more useful than the detailed temporal representation tools. Likewise, it is not clear whether the efficiency in the planning of roads and network infrastructure, especially in the case of pedestrians and cyclists would be significantly increased. This is possibly because data collection is only one part of the process in planning.	
Operability *Related to the ability of the components to keep functioning together*	Record downtime for any of the components along the system pipeline as a percentage of total time	Downtime is minimised to 10% or less
	Progress, Results and Comments:	

[8] https://cvat.org/

	This parameter will be tested at the framework level during R2 period.	
Usability *Related the how well the system helps the users to achieve a task in a given use case*	Interview experts in transport to measure the end-user experience	End-user finds the system easy to use
	Progress, Results and Comments: An online survey to analyse the usefulness and user experience of the tool was mailed to potential respondents. A summary of the general results is given in Table 1, above. In general, the system was found to be interesting and useful by experts. More specific to this use case, the temporal representation of traffic was the only tool that a majority of the respondents did not see being used in the industry.	
Robustness *Related to how robust are the system components during the period of operation*	Sustains performance in various weather conditions, (Power source, sensor, and CPU board operation)	Performance sustained in most weather conditions
	Most GRN hardware has been set up such that weather conditions have minimal effect on performance. On the other hand, RPi based systems overheat quickly. GRN Cameras: the GRN cameras used all have an IP67 rating which means they are waterproof and dustproof, thus are able to sustain performance in many weather conditions. No shutdowns were experienced during the summer months due to excess heat. GRN Edge Layer: The current edge layer is installed indoors, not exposed to any weather elements with air conditioning available if necessary. GRN Fog Layer: The Fog layer is a high-end workstation in a server room so no disruptions due to the weather are expected. In general, the GRN hardware provided is robust.	
Performance *Related to how well the system performs the intended task*	Successfully detect the various events of interest on the road, including trajectories. Availability of the required historical video samples	90% detection of events. 50% detection of trajectories
	Progress, Results and Comments: Events detected by SED and CATFlow yielded the following performance metrics: CATFlow traffic entity per frame detection: mAP@0.5: 0.6927. Trajectories as detected by CATFlow: 84.03% of trajectories are correctly detected averaged over various road scenes. The SED component is used for vehicle detection to detect four vehicle types. A three-fold cross-validation setup was used in the evaluation, and evaluation was done with 1-sec segment-based metrics calculated in a macro-averaged manner. The SED component achieves an F1-score of 50.9% and an error rate of 0.73.	
Accountability *Related to the system being able to explain results or decisions.*	N/A	N/A
	Progress, Results and Comments: This parameter will be tested at the framework level during R2 period	
Transparency *Related to the description of the processes or algorithms that are used to generate system output*	Decision processes are described in a document	Document availability
	Progress, Results and Comments: The processes and algorithms are explained in detail in deliverables D5.4 and D3.2. The former describes the data flow and how the data is transformed as it traverses the E2F2G infrastructure, while the latter describes the AI algorithms that process the data to extract information useful for the use case.	

Privacy awareness *Related to the provision of adequate governance mechanisms that ensure privacy in the use of data.*	Manually evaluate a sample of anonymised AV data to determine privacy protection. Draw a list of secure data characteristics and evaluate each.	Minimise anonymisation misses
	Progress, Results and Comments: Benchmarking results for VideoAnony yielded the following results: the AI model is anonymising videos with a Precision of 0.797, Recall of 0.492, 0.614 mAP@0.5 and 0.29mAP@0.95.	

4.4.2 Use case specific KPIs

This section tabulates the progress in evaluating the use case against a set of functional KPIs (described in Table 6 and evaluated in Table 7) and a set of non-functional parameters (Table 8).

Table 6: Evaluation scenarios and relevant KPIs for the GRN4: Junction Traffic and Trajectory Collection (Table 9, D6.1)

Evaluation Scenario	Target	Relevant KPI
Detection of the trajectories and the storage of these trajectories to be used later in data-driven decision-making.	The aim is to have a 50% detection rate of the trajectories.	**GRN-KPI6 Availability of historical video samples of pedestrian and vehicle trajectories at two junctions.**
Surveys with relevant traffic experts to determine if this data will help in decreasing the planning time.	The aim is to have confirmation that this data is helpful.	**GRN-KPI7 Increased efficiency in the planning of roads**

Table 7: Use case specific KPIs for GRN4: Junction Traffic and Trajectory Collection (Table 6.2 in D1.2)

KPI	Metric	Baseline	Expected result/ Improvement	Evaluators
GRN-KPI6 Availability of historical video samples of pedestrian and vehicle trajectories at two junctions.	Automatic detection of patterns	No baseline	50% of the trajectories detected	GRN managers External traffic experts
	Progress, Results and Comments	Events detected by SED and CATFlow yielded the following performance metrics: CATFlow traffic entity per frame detection: mAP@0.5: 0.6927. Trajectories as detected by CATFlow: 84.03% of trajectories are correctly detected averaged over various road scenes. The SED component achieved an F1-score of 50.9% and an error rate of 0.73.		
GRN-KPI7 Increased efficiency in the planning of roads	Surveys with road planners	No baseline	potential decrease in time for planning through the availability of data	External traffic experts
	Progress, Progress,	An online survey to analyse the usefulness and user experience of the tool was mailed to potential respondents. A summary of the general results is given in Table 5, above.		

KPI	Metric	Baseline	Expected result/ Improvement	Evaluators
	Results and Comments	The survey indicates that the respondents are not confident that the tool will result in a potential decrease in time required in planning and design through the immediate availability of data. On the other hand, one respondent noted that the data collected "can help with noise pollution [studies]". This aspect requires further study.		

Table 8: Use case specific non functional evaluation variables for GRN4: Junction Traffic and Trajectory Collection (Table 6.3 in D1.2)

Evaluation variable	How to measure	Internal evaluators	External Evaluators
End-user experience	Survey	GRN Managers	Transport Experts (Traffic Engineer)
	Progress, Progress, Results and Comments	In general, from the user-experience and usefulness surveys (Table 1, Table 5), the system is rated as satisfactory and can provide useful insights and tools in the transport industry.	
Data protection, privacy preservation	Manually evaluate sample of AV data.	GRN DPO	-
	Progress, Progress, Results and Comments	Benchmarking results for VideoAnony yielded the following results: the AI model is anonymising videos with a precision of 0.797, Recall of 0.492, 0.614 mAP@0.5 and 0.29mAP@0.95. Performance improvement is required on the recall metric since society's demands are high on privacy issues.	
Secure data transmission and cybersecurity in general	Draw a list of secure data characteristics and evaluate each.	GRN Engineers	-
	Progress, Progress, Results and Comments	This parameter will be tested at the framework level during R2 period.	

5 Use case MT1: Monitoring of Crowded Areas

5.1 Introduction

This chapter gathers all the information pertaining to use case MT1: Monitoring of Crowded Areas. The first sections summarise the scope and description of the use case (section 5.2), the components used in the use case (section 5.3), including relevant datasets and the configuration of the framework (section 0). These sections provide a background to the evaluations (relevant to the use case) reported in the sections that follow (i.e., use case specific parameters and KPIs).

5.2 Scope and description of use case

The goal is to select specific cameras (among all those available) that allow monitoring and verifying events like exceptional crowds, suspicious or anomalous behaviour of individuals or unusual crowd movements.

One area for this scenario is a square hosting the “Christmas Markets”, i.e., the fenced area in Piazza Fiera. Every year, from November to the first day of January, in Trento, some of the main squares of the city host the “Christmas Markets” which are visited by thousands of people. Particularly, during the weekend and holidays, these areas are highly crowded. Due to these situations, the number of robberies and aggressions can increase. In addition, first aid may be needed for people who are unwell or faint. A similar area where overcrowding can occur is the square hosting the weekly market located in the city centre, i.e., Piazza Duomo and its vicinities. In addition, this case is more challenging due to the presence of the market stalls' awnings, which invariably occlude camera views.

The MARVEL framework will be deployed to detect the above-mentioned situations or anomalous events that occur in the presence of more than four persons per square meter (or even one person per square meter if COVID-19 restrictions are re-introduced) and alert the operational centre of the local police and consequently the policeman/woman on or close to the site in quasi real-time. In addition to sending alerts, the MARVEL framework will also activate custom views in the control room managed by the local police.

5.3 Components, assets, and framework configuration

Figure 5 gives an overview of the components and frame configuration for the MT1 use case. The data flow starts from the IP cameras (three in Piazza Fiera and another three in Piazza Duomo). Each camera produces continuous video footage. FBK provides the fog tier for all the MT use cases. A secure transmission by VPN access between MT and FBK allows raw video to be sent to the FBK network. In order to comply with the constraints in the agreement that granted FBK access to the raw data of the MT’s sensors and to satisfy the requirements of the MARVdash Kubernetes cluster, FBK deploys two workstations, both with GPU.

Both streams are transmitted to the FBK fog server (MT F1), and then subsequently consumed by the VideoAnony component located at the same server. Upon anonymisation, specifically, by blurring faces present in the received frames, each of the two VideoAnony instances produces an RTSP stream compiled of anonymised frames. These two streams arrive at the second fog device in the MT pilot infrastructure, FBK WS PC (MT F2-non-Kubernetes). This workstation is connected to both the local FBK network (thus enabling communication with the FBK fog server) but also enabled as a Kubernetes node, i.e., as part of the Kubernetes-managed infrastructure. To enable consumption of the VideoAnony streams by the components residing in the Kubernetes-operated part of FBK WS PC (MT F2), an additional streaming service is implemented at the non-Kubernetes part of FBK WS PC (MT F2) – RTSP Proxy service.

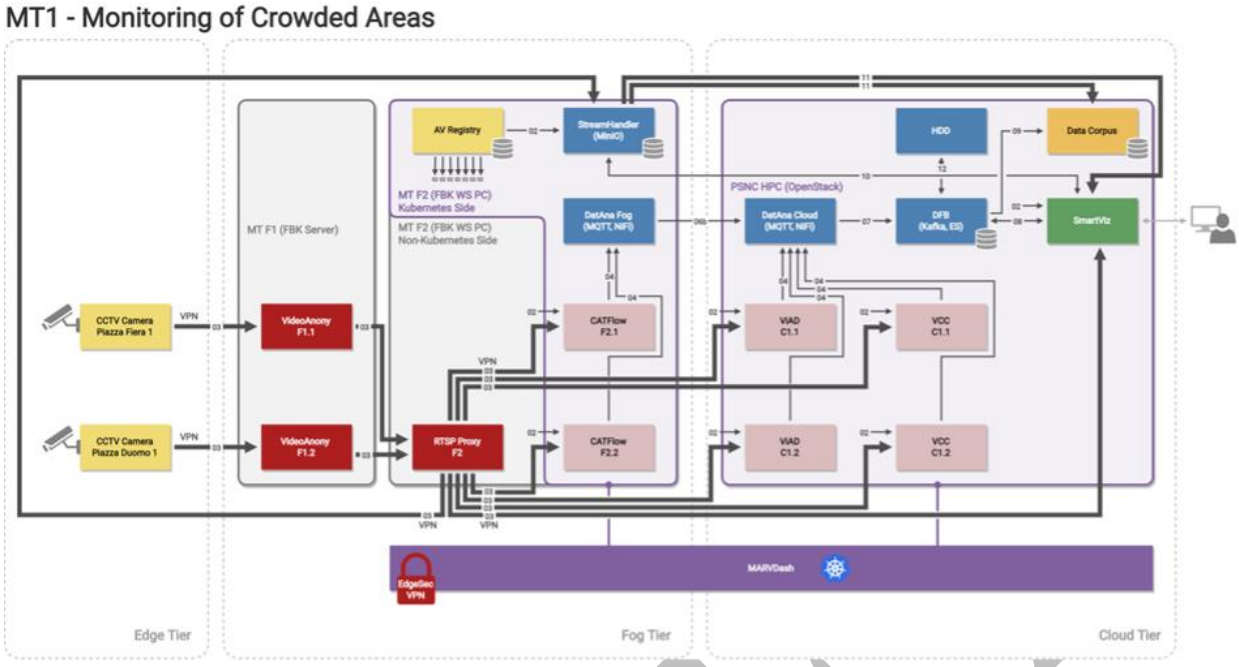

Figure 5: MARVEL R1 deployment and runtime view of the MARVEL architecture for MT1: Monitoring of crowded areas (source: D6.1)

The anonymised streams arriving from the RTSP Proxy service at FBK WS PC (MT F2-non-Kubernetes) are received and processed by CATFlow, ViAD, and VCC, each instantiated two times.

- **CATFlow** – In the MT1 use case, CATFlow was identified to be a good component to track pedestrians for the Piazza Fiera camera. Following a number of tests, it was concluded that CATFlow would also be able to provide the trajectories of these pedestrians. Another camera from this use case at Piazza Duomo is characterised by a less ideal angle and a larger distance from the pedestrians. Despite these challenges, CATFlow can still be used. The vehicle tracking aspect of CATFlow is still employed to track the occasional cyclist passing through or service vehicles that could pass through this pedestrian area.
- **VCC** – The MT1 use case employs a Visual Crowd Counting model to estimate the number of people present in the visual live feed. Crowd counting is the problem of identifying the number of observable people in a visual scene. Information related to the underlying machine learning models developed by MARVEL partners for visual and audio-visual crowd counting can be found in D3.1. For the MT1 use case, the visual crowd counting functionality has not been yet tested on a subset of the MARVEL data since the corresponding model receiving as input visual data has not been trained using this data yet.
- **Visual Anomaly Detection (ViAD)** – The MT1 use case also employs a Visual Anomaly Detection (ViAD) model to identify novel or abnormal events in a visual scene. Information related to the underlying ML model developed by MARVEL partners for detecting anomalies based on visual information can be found in D3.1.

For the development and evaluation of the MT1, a dataset was collected using the video feeds from the selected cameras. The dataset was annotated, anonymised and made available to the partners. The dataset was named “TrentoOutdoor – real recording”, and its details are given in

D2.1[9] and D8.2[10]. The dataset was collated at M11 and will be used by the technical partners to fine-tune, improve and test the AI models. The process will be repeated until the end of the project. At the end of the project, the dataset, or part of it, will be made publicly available to the research community.

Next, the inference results are submitted to the respective MQTT broker of the DatAna agent residing at the same node. Specifically, the CATFlow instances running at FBK fog WS PC edge submit their results to the MQTT broker of DatAna fog agent running at the same fog node – FBK fog WS PC (MT F1-Kubernetes). The two instances of ViAD and VCC submit their results to the MQTT broker of DatAna cloud (PSNC HPC via OpenStack). DatAna fog agent, after receiving the inference results, applies an appropriate transformation to standard data models.

After being gathered in the appropriate Kafka topics in DFB (PSNC HPC via OpenStack), where one Kafka topic per each AI component is configured and enabled, results are consumed by SmartViz and Data Corpus, as detailed below. Finally, DFB passes also results from Kafka topics to Elasticsearch for persistent storage.

Concurrently with processing the RTSP (anonymised) streams for AI inference, another component, StreamHandler, is also subscribed to receive these streams. StreamHandler buffers each of the streams for further segmentation that occurs at regular time intervals, and finally stores the respective AV snippets in the MinIO database. The buffering intervals and segmentation parameters are configurable by the end-user. When a user wishes to inspect the AV data in an interval of interest (e.g., upon detection of an anomaly or raised alert by SmartViz), the user will submit a request to StreamHandler with the onset and offset interval times. StreamHandler then compiles the relevant AV data segments into a single file and sends back to SmartViz the compiled AV data section

The last component of the real-time inference pipeline, also serving as the UI of the platform, is SmartViz. This component runs at the cloud (PSNC HPC via OpenStack). Several functionalities are supported: 1) advanced visualisations, with a dashboard for user configurations and user interactions, 2) live AV feed for real-time inspection of the monitored areas, 3) on-request inspection of stored AV data, and 4) user-based verification of AI inference results.

5.4 KPI tables

This section tabulates the evaluations carried out from the point of view of the use case. These tables are sourced from D1.2 and D6.1 and the evaluations are specific to the use case.

5.4.1 Use Case Parameters

This section tabulates (Table 9) the progress in evaluating the use case against a set of general parameters. Most of these criteria are dependent on the use case KPIs (Table 10) and the asset specific KPIs (Chapter 8).

[9] MARVEL D2.1: Collection and analysis of experimental data, 2021. https://doi.org/10.5281/zenodo.5052713

[10] MARVEL D8.2: MARVEL Data Management Plan, 2021. Confidential.

Table 9: Parameters for use case MT1: Monitoring of Crowded Areas

Parameter	How to measure	Target to be achieved
Efficiency *Related to the efficiency of the system as used in the use case*	Efficiency is largely dependent on how long it takes to detect and flag targeted events. The average time taken to detect and flag the event	On average less time taken when compared to a single person observing multiple cameras
	Progress, Results and Comments: Human operator might not notice relevant events by observing several streams, whereas system performs automatic detection of anomalous event (for R1 anomalous event is the number of people detected in the scene).	
Operability *Related to the ability of the components to keep functioning together*	Install additional cameras and microphones	System discovers sensors or adding sensors does not disrupt operation
	Progress, Results and Comments: Adding a sensor is possible without impacting the system as a whole (or only with limited impacts for the algorithms but this depends on how the AI models are engineered), during the experiments carried out the addition of sensors did not cause problems in terms of system performance. One open issue concerns the bandwidth required by the new sensors and its processing on the fog layer	
Usability *Related the how well the system helps the users to achieve a task in a given use case*	Interview MT staff and local police to measure the end-user experience	End-users finds the system easy to use
	Progress, Results and Comments: Survey exercise will be completed during R2 period.	
Robustness *Related to how robust are the system components during the period of operation*	Scale system to other cameras and microphones	Performance sustained as additional sensors are added
	Progress, Results and Comments: This parameter will be evaluated during the R2 period.	
Performance *Related to how well the system performs the intended task*	Counting the number of detected targeted events in crowds	10% increase in the detection of events when compared to a single person observing multiple cameras
	Progress, Results and Comments: Datasets for CATFlow, VCC and ViAD are fully prepared and utilised for model training and testing. Evaluation will be carried out during R2 period.	
Accountability *Related to the system being able to explain results or decisions.*	System stores video snippet of targeted event	Number of times system fails to store video snippet
	Progress, Results and Comments: In R2, the data storage functionalities (StreamHandler, also MARVEL Data Corpus) will be focused on configurable and event-driven snippet selection. MT staff will check if the system stores video snippets of targeted event and if these will be properly classified.	
Transparency *Related to the description of the processes or algorithms that are used to generate system output*	Decision processes are described in a document.	Document availability
	Progress, Results and Comments: All inference results are stored in DFB and in addition, the end-user is given an explanation of how the system functions. The user will be able to verify each	

Parameter	How to measure	Target to be achieved
	inference result during system operation, which is subsequently captured in the relevant field of the Elasticsearch database.	
Privacy awareness *Related to the provision of adequate governance mechanisms that ensure privacy in the use of data.*	Count data breach reports.	Minimisation of count.
	Progress, Results and Comments: No data breach was reported in RP1. MT staff and FBK staff have complied with all the precepts established by the MT DPO.	

5.4.2 Use case specific KPIs

This section tabulates the progress in evaluating the use case against a set of functional KPIs (described in Table 10 and evaluated in Table 11) and a set of non-functional parameters (Table 12).

Table 10: Evaluation scenarios and relevant KPIs for the MT1: Monitoring of Crowded Areas (Table 11, D6.1)

Evaluation Scenario	Target	Relevant KPI
The evaluation scenario includes testing the various AI models on a labelled dataset to determine the detection rate and score achieved by the models.	Single person observing multiple cameras improve at least 10% the detection of anomalous events.	**MT-KPI1 increase the accuracy in detecting targeted events in crowds**
The evaluation scenario here involves observing the time taken to detect an anomaly through measuring the system's latency.	The aim is to detect anomalies 5 minutes from the start of the anomalous event.	**MT-KPI2 Detection time reduction.**

Table 11: Use case specific KPIs for MT1: Monitoring of Crowded Areas (Table 6.2 in D1.2)

KPI	Metric	Baseline	Expected result/ Improvement	Evaluators
MT-KPI1 increase the accuracy in detecting targeted events in crowds.	Classification accuracy	Single person observing multiple cameras	10% improvement	MT managers
	Progress, Results and Comments	MT staff will check if the system stores video snippets of targeted event and if these are properly classified.		
MT-KPI2 Detection time reduction.	Time	Single person observing multiple cameras		MT managers
	Progress, Results and Comments	MT staff will investigate if the system has allowed the Local Police to intervene in a shorter time than in the period prior to the use of the MARVEL framework.		

Table 12: Use case specific non functional evaluation variables for MT1: Monitoring of Crowded Areas (from Table 6.3 in D1.2)

Evaluation variable	How to measure	Internal evaluators	External Evaluators
End-user experience	Periodic Surveys	MT staff	Local Police
	Progress, Results and Comments	Survey exercise will be completed during R2 period.	
Data protection, privacy preserving	Data breach reports	IT managers/infrastructure managers	DPO
	Progress, Results and Comments	No data breach was reported in RP1. MT staff and FBK staff have complied with all the precepts established by the MT DPO.	
Scalability/Modularity	Extend the solution to other cameras and microphones (other places of city surveillance system)	IT managers/infrastructure managers	Local Police
	Progress, Results and Comments	This parameter is related to operability in Table 9, where progress is reported.	

6 Use case MT3: Monitoring of Parking Places

6.1 Introduction

This chapter gathers all the information pertaining to use case MT3: Monitoring of Parking Places. The first sections summarise the scope and description of the use case (section 6.2), the components used in the use case (section 6.3), including relevant datasets and the configuration of the framework. These sections provide a background to the evaluations (relevant to the use case) reported in the sections that follow (i.e., use case specific parameters and KPIs).

6.2 Scope and description of use case

This use case concerns audio-visual monitoring of a parking lot, including analysis of car trajectories, detection of cars moving out of the parking slots, car damages and robberies, obstructions, etc. The target of this use case is the "Ex Zuffo" Parking Area which is one of the largest parking lots in Trento (around 1000 parking places). It is typically used by citizens who park their cars and then move around the city centre using public transportation, bike-sharing services or e-scooters.

The aim of the use case is to detect anomalies, for example, vandalism and other anomalous behaviour, the timeline distribution of parking activity and the clustering of vehicles or events. When an anomaly is detected, an alert is sent to the Local Police headquarter, who could then check the live feed from the camera where the event is occurring. In addition, the Local Police can also check the feed of a few minutes before the anomaly happened, to accurately assess the cause. The MARVEL framework will therefore support the prevention of robberies or damages to parked cars through audio-video analytics using the existing cameras and the microphones that have been installed thanks to the MARVEL project. The audio-visual analysis will be carried out in real-time and also on recorded data saved on the servers of the Local Police.

6.3 Components, assets, and framework configuration

Figure 6 depicts the frame configuration for the MT3 use case, which in most parts is similar to that deployed in MT1. The main differences lie in the sensing and perception subsystem and in the pool of AI components deployed.

In particular, a microphone is added to the sensing subsystem. The microphone on the ground acquires an audio stream that is passed to the Raspberry Pi (MT E1) and processed by VAD, running on the same device, for detection of speech segments. VAD outputs onset and offset times of the detected speech segments and passes them to AudioAnony to indicate which intervals in the microphone stream require audio anonymisation. AudioAnony anonymises the corresponding audio segments, compiles them together with the remaining audio segments, and finally produces a (single) RTSP stream. The outputs of VAD are submitted to MQTT DatAna brokers residing at the FBK server (MT F1) to be further utilised in the platform (e.g., for audio content inspection by SmartViz).

The video stream is handled as in MT1, i.e., the RTSP stream from VideoAnony is received by an RTSP proxy service residing at the FBK WS PC (MT F2-non-Kubernetes), from where the stream is further forwarded to the AI components for inference. The RTSP stream from AudioAnony travels one more hop than the video counterpart: prior to the RTSP proxy service at the FBK WS PC, this stream goes to an RTSP proxy "relay" located at the FBK server (MT F1), as indicated in the figure. The reason for employing the two RTSP proxy services is to enable data handover between the non-Kubernetes and Kubernetes part of the MT infrastructure.

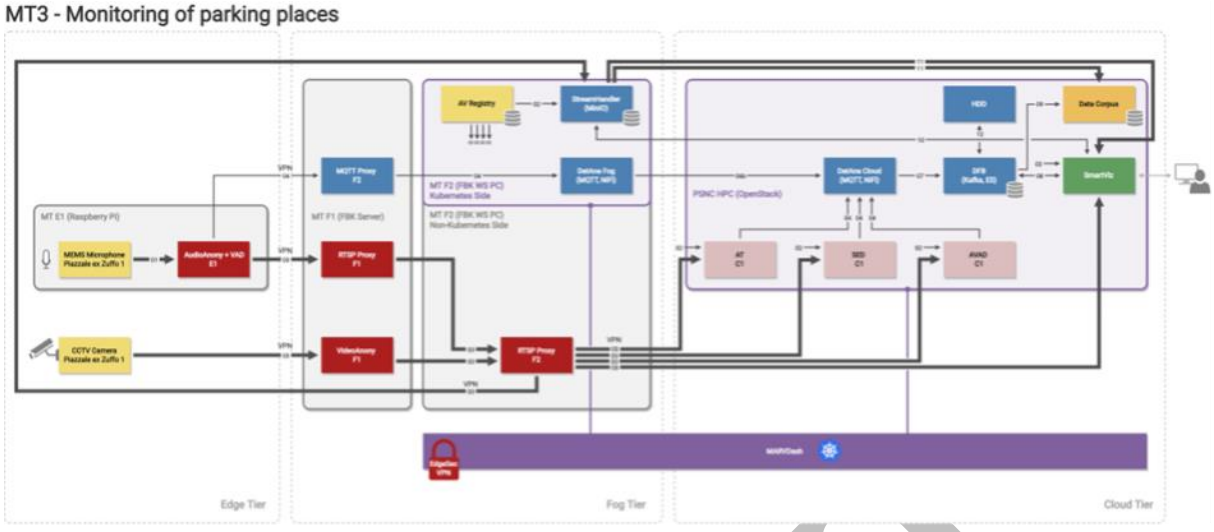

Figure 6: MARVEL R1 deployment and runtime view of the MARVEL architecture for MT3: Monitoring of parking places (source: D6.1)

The anonymised streams arriving from the RTSP Proxy service at FBK WS PC (MT F2-non-Kubernetes) are received and processed by AVAD, SED and AT, each instantiated once, and running at the cloud (PSNC HPC via OpenStack). Following is the pool of AI components deployed in this use case.

- **AVAD** –The component is trained using typical video footages from the Piazzale ex Zuffo parking place, and it recognises any deviation from normal, i.e., events and such that differ from previously seen data.
- **SED** – The component is employed in MT3 for the detection of events from the audio stream relevant to the monitored parking place.
- **AT** – Similarly to SED, the component enables tagging parking audio events at consecutive time intervals.

The outputs from all the AI components are transmitted to the MQTT broker of the cloud DatAna agent (PSNC HPC via OpenStack). The rest of the framework is similar to that deployed in MT1 and the reader is referred to section 5.3.

For the use case, MT has provided the dataset "TrentoOutdoor – real recording" and "TrentoOutdoor – staged recording" which includes video and audio data from cameras and microphones installed at Piazzale ex Zuffo, (details can be found in D2.1 and D8.2).

MT, in collaboration with FBK, collected and anonymised more than 184 videos of 3 minutes each ("TrentoOutdoor – real recording") and thanks to the staged recording done during M13, MT and FBK have provided 38 audio-videos of 30 seconds ("TrentoOutdoor – staged recording"), each manually annotated using ELAN[11].

6.4 KPI tables

This section tabulates the evaluations carried out from the point of view of the use case. These tables are sourced from D1.2 and D6.1 and the evaluations are specific to the use case. Chapters

[11] https://archive.mpi.nl/tla/elan

8 and 9, on the other hand, deal with KPIs that cut across the whole project, i.e., asset KPIs, project KPIs, Business KPIs and societal goals benchmarks.

6.4.1 Use Case Parameters

This section tabulates (Table 13) the progress in evaluating the use case against a set of general parameters. Most of these criteria are dependent on the use case KPIs (Table 14) and the asset specific KPIs (Chapter 8).

Table 13: Parameters to determine for use case MT3: Monitoring of Parking Spaces

Parameter	How to measure	Target to be achieved
Efficiency *Related to the efficiency of the system as used in the use case*	Efficiency is largely dependent on how long it takes to detect and flag dangerous events. The average time taken to detect and flag the event will be measured.	On average less time taken is 5min since onset of dangerous event.
	Progress, Results and Comments: Human operator might not notice relevant events by observing several streams, whereas system performs automatic detection of anomalous event (for R1, anomalous event is car stealing, people argument, bad parking and loud noises). This parameter will be evaluated during the R2 period.	
Operability *Related to the ability of the components to keep functioning together*	Install additional cameras and microphones.	System discovers sensors or adding sensors does not disrupt operation.
	Progress, Results and Comments: The operability parameter is similar to MT1, and progress is reported in Table 9.	
Usability *Related the how well the system helps the users to achieve a task in a given use case*	Interview MT staff and local police to measure the end-user experience via periodic surveys.	End-user finds the system easy to use.
	Survey exercise will be completed during R2 period.	
Robustness *Related to how robust are the system components during the period of operation*	Scale system to other cameras and microphones.	Performance sustained as additional sensors are added
	Progress, Results and Comments: This parameter will be evaluated during the R2 period.	
Performance *Related to how well the system performs the intended task*	Compute accuracy in the detection of targeted events in parking lots .	50% of dangerous events are correctly noted
	Progress, Results and Comments: Dataset for AVAD, SED and AT is fully prepared and utilised for model training. The AT component is used to recognise anomalous scenes in parking places. Evaluation is done with a similar three-fold cross-validation setup. The AT component achieves mAP metric of 0.85. The SED component is used for human action detection in parking places. Evaluation is done with a similar three-fold cross-validation setup. The SED component achieves an F1-score of 58.3% and an error rate of 0.71.	
Accountability	System stores video snippet of targeted event.	Number of times system fails to store video snippet.
	Progress, Results and Comments:	

Related to the system being able to explain results or decisions.	In R2, the data storage functionalities (StreamHandler, also MARVEL Data Corpus) will be focused on configurable and event-driven snippet selection. MT staff will check if the system stores video snippets of targeted event and if these will be properly classified.	
Transparency *Related to the description of the processes or algorithms that are used to generate system output*	Decision processes are described in a document.	Document availability.
	Progress, Results and Comments: All inference results are stored in DFB. For possible future inspection in addition end-user is presented the functionality to verify each inference result during system operation, which is subsequently captured in the relevant field of the Elasticsearch database.	
Privacy awareness *Related to the provision of adequate governance mechanisms that ensure privacy in the use of data.*	Count data breach reports	Minimisation of count.
	Progress, Results and Comments: No data breach was reported in R1. MT staff and FBK staff have complied with all the precepts established by the MT DPO.	

6.4.2 Use case specific KPIs

This section tabulates the progress in evaluating the use case against a set of functional KPIs (described in Table 14 and evaluated in Table 15) and a set of non-functional parameters (Table 16).

Table 14: Evaluation scenarios and relevant KPIs for the MT3, (Table 13, D6.1)

Evaluation Scenario	Target	Relevant KPI
The evaluation scenario involved the increased detection of targeted events, like car damages, checking the number of campers, and the average length of stay.	Single person observing multiple cameras improve at least 50% the detection of anomalous events.	**MT-KPI5 Increase the detection of targeted events**
The evaluation scenario here involves observing the reduction of detection time needed to identify the events mentioned above.	The aim is to detect anomalies 5 minutes from the start of the anomalous event.	**MT-KPI6 Detection time reduction.**

Table 15: Use case specific KPIs for MT3: Monitoring of Parking Spaces (Table 6.2 in D1.2)

KPI	Metric	Baseline	Expected result/ Improvement	Evaluators
MT-KPI5 Increase the detection of targeted events	Classification accuracy	Single person observing multiple cameras	50% of dangerous situations are correctly noted	MT managers
	Progress, Results and Comments	MT staff will check if the system stores video snippets of targeted event and if these will be properly classified.		
	Processing time	Current situation		MT managers

MT-KPI6 Detection time reduction.	**Progress, Results and Comments**	MT staff will investigate if the system has allowed the Local Police to intervene in a shorter time than in the period prior to the usage of the MARVEL framework. The experiment will be planned and executed during the R2 period.

Table 16: Use case specific non functional evaluation variables for MT1: Monitoring of Parking Spaces (from Table 6.3 in D1.2)

Evaluation variable	How to measure	Internal evaluators	External Evaluators
End-user experience	Periodic Surveys	MT staff	Local Police
	Progress, Results and Comments	Survey exercise will be completed during R2 period.	
Data protection, privacy preserving	Data breach reports	IT managers/infrastructure managers	DPO
	Progress, Results and Comments	No data breach was reported in RP1. MT staff and FBK staff have complied with all the precepts established by the MT DPO.	
Scalability/Modularity	Extend the solution to other cameras and microphones (other places of city surveillance system)	IT managers/infrastructure managers	Local Police
	Progress, Results and Comments	This parameter is related to operability in Table 9, where progress is reported.	

7 Use case UNS1: Drone Experiment

7.1 Introduction

This chapter gathers all the information pertaining to use case UNS1: Drone Experiment. The first sections summarise the scope and description of the use case (section 7.2), the components used in the use case (section 7.3), including relevant datasets and the configuration of the framework. These sections provide a background to the evaluations (relevant to the use case) reported in the sections that follow (i.e., use case specific parameters and KPIs).

7.2 Scope and description of use case

The focus of the UNS1 use case is to evaluate MARVEL technologies using drones in controlled environments and to support partners' use cases for the development and validation of the federated learning framework. Such technologies could help society in the task of monitoring and surveillance of large public events as there could be unpredictable behaviour of the crowd and frequently there is a lack of fixed infrastructure for monitoring. Also, fixed street cameras can provide frontal views of the crowd, but inner details could not be checked accurately due to occlusions. Furthermore, some viewing angles or even whole spaces are not covered using fixed cameras.

The utilisation of drones equipped with cameras and computational resources, complemented with additional ground microphones, could help to quickly check whether some problematic behaviour has occurred among the crowds. As the drone flies above the main event points, video recordings could help in recognising if an anomalous and potentially dangerous event is happening. If the camera on the drone spots a problem, the drone can move closer or inform the security of the problem. The use of the drone coupled with anomaly detection algorithms is therefore expected to reduce the time taken for security personnel to detect problems since their onset.

The focus of the use case is also on federated learning for crowd counting and onboard real-time processing. The idea of the pilot is to perform crowd classification and crowd counting. Regarding classification, we will focus on the distinction between three classes of crowd behaviour: Neutral, Party, and anomalous/dangerous behaviour. Later on, the classes can be further refined/multiplied as needed.

UNS members had several meetings with the organisers of the EXIT festival, which is one of the largest open-space music events in Europe, held each summer in Novi Sad. Their feedback was used as guidance for upgrading evaluation scenarios, as they are a potential end-user. As a result, crowd counting was identified as the most desirable task to work on and their feedback was used while performing staged recordings.

7.3 Components, assets, and framework configuration

The UNS inference pipeline consists of 3 layers – Edge, Fog and Cloud, as shown in Figure 7. Data capturing is performed at the edge using the AVDrone component. The sensors consist of a GoPro camera mounted on the drone and ground-based IFAG AudioHub MEMS microphones. The GoPro camera is connected to an Intel NUC, whereas microphones are connected to Raspberry Pi v4 boards on the ground.

Video anonymisation and audio anonymisation are enabled by installing VideoAnony on the same Intel NUC and AudioAnony on the same Raspberry Pi boards, respectively. Data streaming is further enabled via RTSP. AudioAnony is coupled with VAD within the same container, whose purpose is to detect voice activity boundaries, which are further provided to

AudioAnony for controlling the activation of the anonymisation process. The outputs of VAD are submitted to MQTT DatAna brokers, also residing at the Raspberry Pi, to be further utilised in the platform (e.g., for audio content inspection by SmartViz). The anonymised data is further transmitted to the fog tier where AI models are executed.

VCC is chosen as one of the AI components of interest within R1. Besides VCC, FedL-based VCC model training is implemented at the Fog. VCC-FedL client runs on a GPU installed on the UNS Fog server and it also communicates with the VCC-FedL server that runs on the cloud (PSNC HPC via OpenStack). SED is another inference component implemented at the fog. The other components at the fog tier include DatAna Fog (MQTT, NiFi), AV Registry and StreamHandler (MinIO). At the cloud tier the following components are implemented: DatAna Cloud, DFB, HDD, SmartViz, and DataCorpus.

Anonymised data is then transferred to the fog tier, to perform inference using VCC and to inspect streams using SmartViz. These two components and StreamHandler request metadata details of the AudioAnony and VideoAnony within the initialisation phase from AV Registry, deployed at a server. If the inference output of the VCC component is higher than the set threshold, an alarm is raised so that a human operator can observe the snippet of interest in the stream.

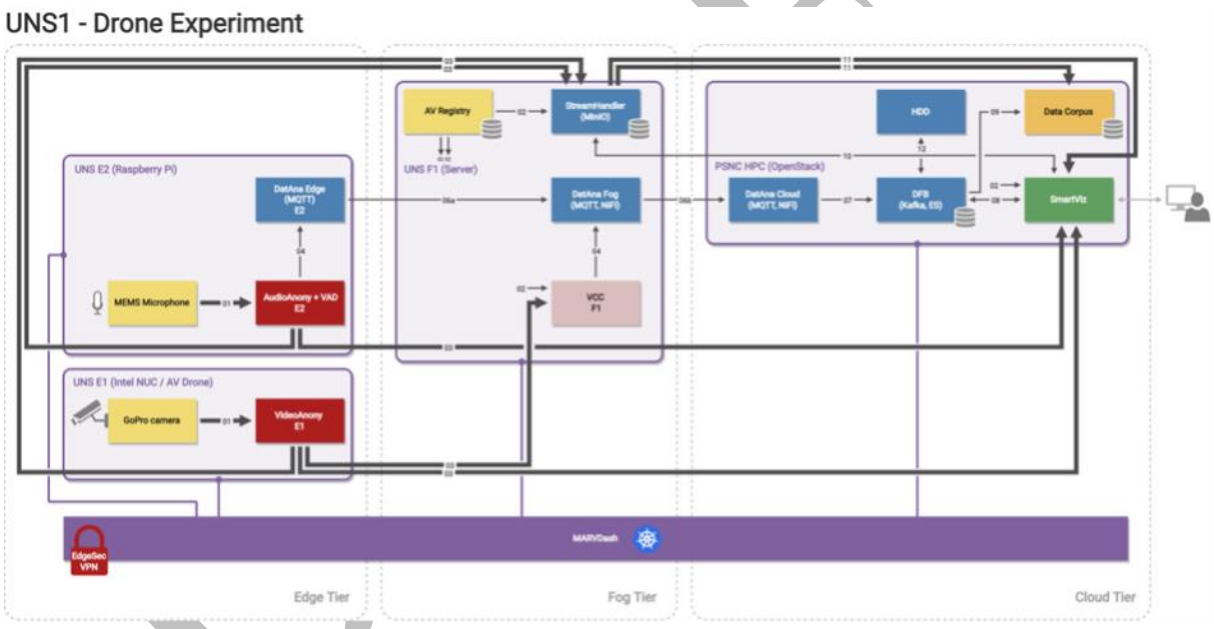

Figure 7: MARVEL R1 deployment and runtime view of the MARVEL architecture for UNS1: Drone Experiment (source: D6.1)

For AI component training, UNS provided the "UNS Drone dataset". It consists of 55 minutes of raw video snippets, making a total of 11.4GB of video data. Data is recorded using HD resolution, 30fps and H.264 standard. Raw video data is further annotated for visual crowd counting and 819 annotated frames are provided. Annotations were made using the CVAT tool and they are exported in xml format. Besides video data, audio data is recorded using IFAG AudioHub Nano microphones and we will deal with audio annotations within R2 in order to prepare the audio-visual dataset. Audio-visual data were recorded within the staged recording process and recordings contain captures of various sound or visual anomalies in the crowd, including the simulated sound of gunshots, and broken glass causing people to disperse/run away. The above-mentioned dataset features snippets capturing variations in several aspects: distances, camera angles and number of people in the scene. The data were acquired during

UNS staged recordings in M16 at the Petrovaradin fortress, where UNS staff and students were gathered to simulate crowd movements and relevant crowd events. This dataset contributes to the *UNS drone dataset* described in detail in D2.1 (Section 4.2.6) and also D1.2 (Section 4.6.1) and D8.2. Due to the drone recording limitations (specifically, the light angle needs to satisfy certain requirements that prevent early recordings), the recordings had to occur during times of day where the likelihood of accidental observers increases. We note that all experimental participants have signed written consent to participate in the staged recordings.

7.4 KPI tables

This section tabulates the evaluations carried out from the point of view of the use case. These tables are sourced from D1.2 and D6.1 and the evaluations are specific to the use case.

7.4.1 Use Case Parameters

This section tabulates (Table 17) the progress in evaluating the use case against a set of general parameters. Most of these criteria are dependent on the use case KPIs (Table 18) and the asset specific KPIs (Chapter 8).

Table 17: Parameters to determine for use case UNS1: Drone Experiment

Parameter	How to measure	Target to be achieved
Efficiency *Related to the efficiency of the system as used in the use case*	Efficiency is largely dependent on how long it takes to detect events. The average time taken to detect and flag the event is measured.	To decrease the time needed to identify an event using audio-visual monitoring as compared to human visual detection (by security crew).
	Progress, Results and Comments: Human operator might not notice relevant events by observing several streams, whereas system performs automatic detection of anomalous event (for RP1 anomalous event is number of people higher than a threshold). The parameter will be experimentally evaluated during R2 period.	
Operability *Related to the ability of the components to keep functioning together*	Add additional multi-modal input sources.	Different data modalities successfully accommodated in the platform (audio, video, GPS, etc.).
	Progress, Results and Comments: Audio and video input data sources are successfully implemented and both data sources are supported currently.	
Usability *Related to how well the system helps the users to achieve a task in a given use case*	Interview security crew of public event organisers to measure the end-user experience via periodic surveys.	End-user finds the system easy to use.
	Progress, Results and Comments: Within UNS Drone Experiment, SmartViz provides a user-friendly interface. For this use case, SmartViz provides three main functionalities: • to raise an automatic alarm; • to play live AV data feed; • to request historical AV data, to check how the crowd gathered. An automatic alarm would be raised if there is a larger number of people in the crowds than a pre-defined threshold. This is supported using VCC component in the backend. Besides VCC, VAD is also supported in SmartViz. This way, a human operator can focus on anomalies in the crowds. Inspection of several data streams would be much more difficult without the help of developed framework.	

Parameter	How to measure	Target to be achieved
	For the next release, visualisation inference results and data streams could be improved and we will work towards splitting screen into several parts due to easier inspection. Besides that, VAD could be visualised by showing decibel level of recorded speech and by using different colours.	
Robustness *Related to how robust are the system components during the period of operation*	Change operating conditions (namely distance and light intensity) and measure accuracy, latency, and packet loss.	5% improvement on current operating conditions
	Progress, Results and Comments: We will explore experimental setup that includes ping measurement in order to check connectivity details between two tiers.	
Performance *Related to how well the system performs the intended task*	Classification accuracy across multimodal sources.	To increase the average accuracy (5%) for the drone-based audio-visual anomaly detection as compared to baseline (vision only).
	Progress, Results and Comments: Dataset for VCC is fully prepared and utilised for model training. Raw dataset for AVCC is fully prepared and evaluation will proceed upon audio annotations.	
Accountability *Related to the system being able to explain results or decisions.*	System stores video snippet of event detected.	Number of times system fails to store video snippet.
	Progress, Results and Comments: In RP2 the data storage functionalities (StreamHandler, also MARVEL Data Corpus) will be focused on configurable and event-driven snippet selection. The proposed experimental setup to evaluate this KPI is the following: Within a staged recording process, a concrete number of events is simulated at a certain point in time and then compared to the number of stored video snippets and timestamps of event detected in order to find out the number of times systems fails.	
Transparency *Related to the description of the processes or algorithms that are used to generate system output*	Decision processes are described in a document.	Document availability.
	Progress, Results and Comments: All inference results are stored in Elasticsearch (DFB). For possible future inspection in addition end-user is presented with the functionality to verify each inference result during system operation, which is subsequently captured in the relevant field of the Elasticsearch database.	
Privacy awareness *Related to the provision of adequate governance mechanisms that ensure privacy in the use of data.*	Periodically evaluate the secure transmission.	Minimise data breaches.
	Progress, Results and Comments: Anonymisation is performed at devices that capture raw data (Video anonymisation at drone onboard Intel NUC, Audio anonymisation in the field Raspberry Pi). Data transmission security is enabled via EdgeSec VPN.	

7.4.2 Use case specific KPIs

This section tabulates the progress in evaluating the use case against a set of functional KPIs (described in Table 18 and evaluated in Table 19) and a set of non-functional parameters (Table 20).

Table 18: Evaluation scenario and relevant KPIs for the UNS1, (Table 15, D6.1)

Evaluation Scenario	Target	Relevant KPI
Detection of anomalous events and alerting event organisers about them,	The goal is to achieve 5% improvement comparing to the frontal camera views only or vision only.	**UNS-KPI1 Increase the average accuracy for the drone-based audio-visual anomaly detection.**
System performance evaluation against, e.g., accuracy, latency, packet loss, for varying operating system conditions	The goal is to achieve at most 10% performance degradation comparing to the nominal system operation	**UNS-KPI2 Robustness to different operating conditions (e.g., distance, fps rate, camera resolution, modality dropout)**
Monitoring (e.g., by security crew) of streaming data.	An improvement of 5% in detection time is expected in comparison to human-based monitoring.	**UNS-KPI3 Decrease the time needed to identify an event using audio-visual monitoring comparing to the human (manual) detection.**
System performance evaluation with presence of both audio and visual modality and modality dropout.	The goal is to achieve 5% improvement comparing to the video or audio-based only system operation.	**UNS-KPI4 Multimodality: Different data modalities successfully accommodated in the platform (audio, video, GPS, etc.)**

Table 19: Use case specific KPIs for UNS1: Drone Experiment (Table 6.2 in D1.2)

KPI	Metric	Baseline	Expected result/ Improvement	Evaluators
UNS-KPI1 Increase the average accuracy for the drone-based audio-visual anomaly detection	Classification accuracy; Multimodality	Vision only case	5% improvement	UNS staff
	Progress, Results and Comments	Dataset for VCC is fully prepared and utilised in model training. Raw dataset for AVCC is fully prepared and evaluation will proceed upon audio annotations, during the R2 Period.		
UNS-KPI2 Robustness to different operating conditions (e.g., distance, fps rate, camera resolution, modality dropout)	Accuracy (Light intensity) Latency (distance) Packet loss (distance)	Current System operation	At most 10% performance degradation compared to the nominal system operation	UNS staff
	Progress, Results and Comments	We will explore experimental setup that includes ping measurement in order to check connectivity details between two tiers.		
UNS-KPI3 Decrease the time needed to identify an event using audio-visual monitoring	Processing time	Human visual detection (e.g., by security crew)	Improvement of 5% in detection time is expected in comparison to human-based monitoring.	UNS staff

comparing to the human (manual) detection	**Progress, Results and Comments**	SmartViz supports the raise of automatic alarm if there is a larger number of people in the crowds than a pre-defined threshold. This way, a human operator can focus on anomalies in the crowds, whereas inspection of several data streams at the same time would be much more difficult without the help of developed framework.		
UNS-KPI4 Multimodality: Different data modalities successfully accommodated in the platform (audio, video, GPS, etc.)	Classification accuracy	Video-only or audio-only system operation	The goal is to achieve 5% improvement compared to the video or audio-based only system operation	UNS staff
	Progress, Results and Comments	In the event of video modality dropout, VAD utilised to detect human presence.		

Table 20: Use case specific non functional evaluation variables for UNS1: Drone Experiment (6.3 in D1.2)

Evaluation variable	How to measure	Internal evaluators	External Evaluators
Modularity	Integration of new equipment	Infrastructure managers	IT experts of security crews
	Progress, Progress, Results and Comments	Modularity of the framework is supported two-fold: in terms of computing devices and in terms of data capturing devices. With regards to the computing devices, modularity is supported through MARVdash components wherein adding a new device is achieved by installing appropriate Kubernetes tools and by associating the device to the relevant Kubernetes cluster. Modularity in terms of sensing devices is achieved by updating the AVRegistry with appropriate device metadata and withdrawing this information from AI components that should process the corresponding data streams. The mechanisms for automatic AI component deployment configuration and source/stream assignment are planned for RP2.	
Secure transmission	Periodic evaluations	UNS staff	IT experts of security crews
	Progress, Progress, Results and Comments	Planned for RP2	
End-user experience	Periodic surveys	UNS staff	Security crew of public events organisers (e.g., city authorities)
	Progress, Progress, Results and Comments	At the current stage UNS academic staff was performing evaluation of the results. Within UNS Drone Experiment, SmartViz provides a user-friendly interface. For this use case SmartViz provides three main functionalities: • to raise an automatic alarm; • to play live AV data feed; • to request historical AV data, to check how the crowd gathered. An automatic alarm would be triggered if there are a larger number of people in the crowds than a pre-defined threshold. This is supported using VCC component in the backend. Besides VCC, VAD is also supported in SmartViz. This way, a human operator can focus on anomalies in the crowds. Inspection of several data streams would be much more difficult without the help of developed framework.	

8 Asset Specific KPIs

This chapter deals with the evaluation parameters and KPIs that are asset specific. These component-based KPIs are not use case specific, but cut across a number of use cases, depending on whether they are used or not used in each specific use case.

8.1 Sensing and perception subsystem

Table 21: Asset specific KPIs for the "Sensing and perception subsystem"

Asset	KPI	Metric	Baseline SOTA	Dataset/ Benchmarks	Expected result	Relevant Project KPIs
IFAG-MEMS	Performance	Distortion Sensitivity Phase tolerances Frequency roll-off Multichannel synchronisation	Current release of the microphone	MARVEL dataset	High performance	KPI-O1-E1-2
	Progress, Progress, Results and Comments	Acquisition boards with up to 4 microphones have been developed to assure the synchronisation among channels. All boards implement SOTA IFAG microphones, which present high SNR, low distortion, small phase tolerance and frequency roll-off. A board with 8 microphones is being developed, adding on top of the channel increase also a microprocessor more capable to perform audio pre-processing.				
	Robustness	Usability on drones, intersections, and public spaces	various possible uses	-	Can be implemented in all MARVEL use cases	
	Progress, Progress, Results and Comments	The acquisition boards developed are being used in public spaces scenarios. The new board on develop counts with a housing water/dust proof, making it suitable for outdoor scenarios.				
GRNEdge	Robustness	Sustains performance in various weather conditions, (Power source, sensor, and CPU board operation)	Experiments during MARVEL project	GRN AV data from field trials	Performance sustained in most weather conditions	KPI-O1-E1-1 KPI-O1-E1-2 KPI-O5-E1-1
	Progress, Progress, Results and Comments	Two devices were provided by GRN to serve as the GRNEdge: the GRNEdge V1 and The GRN Edge PC. The GRNEdge V1 device had enough memory and power to run for about 2 hours. The device was placed in a weatherproof case that is able to withstand rain and wind. The device was tested in real-world conditions during the month of August, where temperatures are 22-30ºC. These conditions, as well as the inability to have active cooling in the transparent box, made it difficult to use the device for periods of time longer than 15 minutes. The GRN Edge PC can be connected to the GRN IP cameras and still be placed indoors. The GRN IP Cameras are rated at IP67 thus can be considered to be waterproof and dustproof. The GRN Edge PC is protected from all elements and provided with air conditioning if necessary. Thus, can be considered to be robust to all weather conditions.				
	Reliability	Percentage of transmitted packets lost	Experiments during	GRN AV data from field trials	Downtime is minimised to 10% or less	

Asset	KPI	Metric	Baseline SOTA	Dataset/ Benchmarks	Expected result	Relevant Project KPIs
			MARVEL project			
	Progress, Progress, Results and Comments	GRNEdge V1 portable device performs as intended, recording video for a short time. However, the difficulties with overheating due to the weatherproofing case used to protect the device make it improbable that a similar setup can be used for long-term data collection without having considerable downtime. The GRN Edge PC has only experienced downtime in cases of power outages which are rare occurrences. Thus, the goal of minimising downtime has already been reached.				
SED@edge	Classification accuracy	Classification Top-1 Accuracy	SOTA for SED on the edge SOTA SED methods	UrbanSound8k, ESC50, MARVEL use cases annotated data	Less than 5% accuracy drop for RPi model, less than 10% accuracy drop for STM32 implementation.	KPI-O2-E2-1 KPI-O2-E2-2 KPI-O2-E2-3
	Progress, Progress, Results and Comments	Target model reduction reached. Accuracy drop is still to be improved.				
	Computational complexity	MAC count			Enabling real-time event-detection: depends on the platform	
	Progress, Progress, Results and Comments	The current version of SED@Edge achieves real-time processing requirements on MCUs.				
	RAM occupation	RAM occupancy			Small enough to be run on a STM32 MCU (~300KB)	
	Progress, Progress, Results and Comments	The current version of SED@Edge can run on the RAM of an STM32 MCU.				
	Energy efficiency	Power consumption			In the mW range, enough for solar harvester	
	Progress, Progress, Results and Comments	The current version of SED@Edge achieves the power constraints requirements.				
AVDrone	Streaming data and real or near real-time decision-making	Latency, Robustness, Data-transfer rate	Visual human inspection; decision-making based on visual only data	UNS drone experiment	Real-time or near real-time decision-making at the edge or at the fog, achieved by different edge processing units and/or by employing	KPI-O1-E2-2 KPI-O1-E2-3 iKPI-12.2

Asset	KPI	Metric	Baseline SOTA	Dataset/ Benchmarks	Expected result	Relevant Project KPIs
					compressed ML/DL models	
	Progress, Progress, Results and Comments	Visual human inspection is provided at the fog tier by streaming AV data to the laptop at the fog. Data transfer rate that supports HD video stream is supported with a low latency. Near real-time inspection at the fog tier is supported.				
	Robustness to different operating conditions (e.g., horizontal and vertical distance to the event of interest, varying payload)	Coverage distance, Hovering time	Single wireless connectivity; small payload (no edge processing unit)	UNS drone experiment	Two wireless connectivity options (Wi-Fi and, e.g., LTE-M) enabled; hovering time of 10-15 minutes with edge processing payload (tentative)	
	Progress, Progress, Results and Comments	Hovering time is constrained due to the capacity of batteries and it is about 15 minutes long with payload that includes INTEL N|UC and camera. Theoretically, maximal transmission distance of DJI M600 is up to 5km, but coverage distance strongly depends on the available network connection. Horizontal and vertical distance to the event of interest depends on drone types classification and laws in a country of interest.				
SensMiner	Amount of collected data	Data amount in Giga Bytes	Audio data collected from drone only	UNS drone experiments	Audio data increased by 10%	KPI-O1-E1-2
	Progress, Progress, Results and Comments	The application has not been exploited yet.				
	Usefulness	number of useful audio events being captured	Audio data collected from drone only	UNS drone experiments	Number of useful (annotated) audio events increased by 10%	
	Progress, Progress, Results and Comments	The application has not been exploited yet.				

8.2 Security, privacy and data protection

Table 22: Asset specific KPIs for the "Security, privacy, and data protection" subsystem (6.6 in D1.2)

Asset	KPI	Metric	Baseline SOTA	Dataset / Benchmarks	Expected result	Relevant Project KPIs
EdgeSec	Usability	Effort needed by end-user	Other commonly deployed VPN-based solutions	N/A	Least possible manual tuning from the component users	KPI-O1-E3-2 KPI-O3-E3-1 iKPI-2.2

Asset	KPI	Metric	Baseline SOTA	Dataset / Benchmarks	Expected result	Relevant Project KPIs
	Progress, Progress, Results and Comments	**EdgeSec VPN** is containerised thus, the effort required for the setup of the component is reduced to just running the container. The container image itself does not have any dependencies and does not require advanced knowledge of networking systems. The end-user simply provides an IP address and port of the Super Node. More details can be found in D4.2[12]. **EdgeSec TEE** will contribute to the fulfilment of the above KPIs. However, since it is not fully integrated with the MARVEL platform, the experimental results have not been explored yet.				
	Scalability	Network performance metrics (latency)	N/A	N/A	Zero or barely noticeable increase in network traffic latency due to the deployment of EdgeSec	
	Progress, Progress, Results and Comments	By conducting measurements of the Round-trip time of an ICMP packet with and without the presence of **EdgeSec VPN,** we can examine the imprint of the component to the network. The reported overhead for packet sizes of 100, 500, and 1000 bytes was 0.571ms, 0.551ms, and 0.526ms respectively on average. Those numbers indicate plenty of room for scalability. More details can be found in D4.2. **EdgeSec TEE** will contribute to the fulfilment of the above KPIs. However, since it is not fully integrated with the MARVEL platform, the experimental results have not been explored yet.				
	Effectiveness for avoiding Cyber threats	Number of threats avoided	N/A	N/A	3 distinct cyber threats avoided	
	Progress, Results and Comments	**EdgeSec VPN**, like any VPN solution, mitigates attacks such as ISP snooping, attacks over insecure wireless networks as well as compromised networking equipment. More details can be found in D4.2. **EdgeSec TEE** will contribute to the fulfilment of the above KPIs. However, since it is not fully integrated with the MARVEL platform, the experimental results have not been explored yet.				
	Communication Security	amount of encrypted data in transit	N/A	N/A	100% of data will be end-to-end encrypted	
	Progress, Results and Comments	**EdgeSec VPN** has four ciphers built-in namely Twofish, AES, ChaCha20, and SPECK. Any of these ciphers can be used to encrypt data in transit. We verified that data in transit is encrypted with a simple tcpdump command. More specifically with a simple python-based HTTP server we request the contents of the webpage with and without the presence of EdgeSec VPN. The tcpdump command shows the contents of the webpage in clear text without the presence of EdgeSec VPN. On the other hand, when EdgeSec VPN is present, tcpdump shows scrambled data proving that data is encrypted. More details can be found in D4.2. **EdgeSec TEE** will contribute to the fulfilment of the above KPIs. However, since it is not fully integrated with the MARVEL platform, the experimental results have not been explored yet.				
Video Anony	Image naturalness	Frechet Inception Distance (FID)	Classic techniques, e.g., blurring	MOT, CelebA, and annotated data from the	A lower FID value than the baselines	KPI-O1-E3-1 KPI-O1-E3-3

[12] MARVEL D4.2: Security assurance and acceleration in E2F2C framework – initial version, 2022. https://doi.org/10.5281/zenodo.6821254

Asset	KPI	Metric	Baseline SOTA	Dataset / Benchmarks	Expected result	Relevant Project KPIs
	Face detection	Percentage of detected faces	and pixelation, and some recent state-of-the-art GAN-based face conversion techniques, e.g., DeepPrivacy and CIAGAN	MARVEL use cases	Higher than baselines	
	Face de-identification	Percentage of correctly identified faces			Lower or comparable than GAN-based baselines	
	Real-time performance	Time in seconds			At least 10% reduction compared to the uncompressed model	
	Progress, Results and Comments	Performance goals are under evaluation, preliminary results indicate that they are achieved. Model reduction achieved by replacing convolutions with depth-wise separable convolutions and via quantisation.				
Audio Anony	Voice anonymisation	EER	Voice activity detection, signal processing voice anonymisation method	Librispeech or VoxCeleb data, and MARVEL annotated corpora	At least 50% EER	KPI-O1-E3-1 KPI-O1-E3-3
	Amount of distortion	Signal-to-distortion ratio, focusing on the non-speech sound events			20% PESQ improvement over baseline. 20% SED improvement over VAD or baseline	
	Computational complexity	RAM, MOPS	Baseline of the Voice privacy challenge		10% complexity reduction	
	Progress, Results and Comments	Performance goals are being monitored. PESQ target is achieved. EER is a bit lower 46% on VTCK. The solution is still under development and therefore no final evaluation is available. Reduction still to be addressed				
devAIce	Voice detection	F1	Models deployed in devAIce	Artificially mixed dataset MARVEL corpus, public dataset (e.g., LibriSpeech, and AudioSet), eventually contaminated	10% improvement	KPI-O1-E1-2
	Progress, Results and Comments	The F1-score increased from 0.77 with the old model to 0.86 with the new model.				
	Computational complexity	RAM and CPU usage	Models deployed in devAIce	MARVEL corpus, public dataset (e.g., LibriSpeech, and AudioSet), eventually contaminated	10% improvement	

Asset	KPI	Metric	Baseline SOTA	Dataset / Benchmarks	Expected result	Relevant Project KPIs
	Progress, Results and Comments	The time elapsed on an input of 1449s on a multi-core CPU decreased from 5.44s with the old model to 1.37s with the new model. RAM usage is minor for both models (<20MB).				
	Robustness	Precision, recall	Models deployed in devAIce	MARVEL corpus, public dataset (e.g., LibriSpeech, and AudioSet), eventually contaminated	10% improvement	
	Progress, Results and Comments	Recall increased from 0.74 to 0.82 - Precision increased from 0.82 to 0.92.				

8.3 Data management and Distribution

Table 23: Asset specific KPIs for the "Data management and Distribution"

Asset	KPI	Metric	Baseline SOTA	Dataset / Benchmarks	Expected result	Relevant Project KPIs
DFB	Data Integrity	Data loss rate	Isolated execution of DFB	Synthetic data streams	Confirm that advanced encryption mechanisms over end-to-end data transfer will guarantee data integrity	KPI-O1-E2-1 KPI-O1-E2-2 KPI-O1-E2-3 KPI-O1-E2-4
	Progress, Results and Comments	Data loss rate: 0 (MVP release, as reported in D5.2 and R1 release).				
	Scalability	Hardware resources utilisation, speedup			Increase the number of modality data streams and verify that performance metrics improve or at least stay the same	
	Progress, Results and Comments	R1 supported modality data streams: 10 (corresponding to the available distinct streams/topics for inference results of AI components: CATFlow-V, CATFlow-P, TAD, ViAD, AVAD, VCC, AVCC, SED, AT, VAD). Hardware speedup measurements for R1 are currently in progress in accordance with D5.5 guidelines. The plan is to measure latency and throughput under different load scenarios.				
	Availability	Service availability-failed request, data access restriction			Verify that DFB resources are available and discoverable	
	Progress, Results and Comments	Service availability:100% Data access restriction: None (MVP release, as reported in D5.2 and R1 release)				
	Performance for high volume, heterogeneous data streams	Data transfer latency, data throughput, response time, number of cluster nodes			Thoroughly measure different performance metrics under different execution conditions	

Asset	KPI	Metric	Baseline SOTA	Dataset / Benchmarks	Expected result	Relevant Project KPIs
	Progress, Results and Comments	Data transfer latency: 5 ms (200 MB/s load) Data throughput: 605 MB/s Response time: 5 ms (200 MB/s load) Number of cluster nodes:3 Measurements are from the MVP release, as reported in D5.2. Respective measurements from R1 are in progress.				
DatAna	Performance in data rates	Amount of data processed and events per second -per node, typically per processor	Isolated execution in layers not interconnected	Interconnection of NiFi topologies to enable inference data communication between edge/fog and cloud.	Improve the overall performance of the data flows managed by DatAna by creating template data flows for the most typical scenarios in data acquisition/processing using NiFi off-the-shelf processors and enabling the NiFi Site-toSite (S2S) communication to move inference data in a secure and efficient way	KPI-O1-E2-1 KPI-O1-E2-2 KPI-O1-E2-3 KPI-O1-E2-4
	Scalability	Horizontal scalability in a NiFi cluster, MiNiFi agents handled in multiple devices	NiFi capabilities for horizontal scalability up to 1000 nodes	In MARVEL there is no need to apply horizontal scalability so far. NiFi deployed at the Cloud layer not in cluster mode	Allowing the creation of topologies of Apache MiNiFi/NiFi instances in fog to cloud scenarios. NiFi is deployed at the Cloud, MT, GRN and UNS Fog and at GRN Edge. This load balance the traffic and the data throughput among the different layers and pilots.	
	Progress, Results and Comments	Time to process and transform an entry in NiFi: 0,06ms average 100% availability 0 data loss rate Measurements to be done for R1.				
HDD	Data management and distribution	Data access latency, data loss	Purely centralised or distributed alternative data distribution schemes	MARVEL-tailored simulation environments	Execution time of data management and distribution improved at least 15%. For example, on data access latency, via breaking down the distribution process in different segments, such as data moving time, data processing time, computation time, and inter-node delays	KPI-O1-E2-1

Asset	KPI	Metric	Baseline SOTA	Dataset / Benchmarks	Expected result	Relevant Project KPIs
	Progress, Results and Comments	Significant performance gains across a variety of metrics against simulated industrial recommendations from Microsoft and Confluent.				
	Device resource usage	Network lifetime, energy consumption rate and energy balance, number of special nodes/equipment	Purely centralised or distributed alternative data distributio n schemes	MARVEL-tailored simulation environments	Improve data distribution in relevant device resource usage by at least 15%. For example, on energy consumption, processing time, and memory allocation, via decentralised approaches and low complexity algorithms, as well as targeting the related resource consumption vs. accuracy trade-offs.	KPI-O1-E2-4
	Progress, Results and Comments	Significant performance gains across a variety of metrics against simulated industrial recommendations from Microsoft and Confluent.				
Stream Handler	Performance for high volume, heterogeneous data stream	Data transfer latency, data throughput, response time	Isolated execution of StreamHan dler	Synthetic audio-visual streams of different formats and quality	Ensure small latency in - storage of segments, - response time to API requests (including compilation of edited files). Validate with streams of different formats and degrees of quality	KPI-O1-E2-2
	Progress, Results and Comments	The introduction of StreamHandler's new functionality clearly increases the number of modalities handled by the project's Data Management Platform (DMP) as these now include large files of video and audio data. Thus, the KPI can be considered as already achieved. In terms of latency and response times to API requests, all measurements so far show extremely quick responses that are actually bound by the infrastructure at each case and mainly by the performance of the attached storage. The validation with streams of different formats and degrees of quality, in terms of bitrate and resolution, is expected to produce concrete results in the immediately following period and will be reported in D5.5.				

8.4 Audio visual and multimodal AI

Table 24: Asset specific KPIs for the "Audio visual and multimodal AI" subsystem (Table 6.8 in D1.2)

Asset	KPI	Metric	Baseline SOTA	Datasets / Benchmarks	Expected result	Relevant Project KPIs
AVAD	Accuracy	Area Under ROC Curve	Multi-Modal Anomaly Detection by Using Audio and Visual Cues	No relevant public dataset available. EMO&LY, synthetic dataset is of a related task We created a synthetic dataset based on the dataset	95% on synthetic dataset	KPI-O2-E2-1 KPI-O2-E3-1 KPI-O2-E3-2 KPI-O2-E3-3 iKPI-3-3
	Speed	FLOPS			30% speedup while retaining 90% accuracy	

Asset	KPI	Metric	Baseline SOTA	Datasets / Benchmarks	Expected result	Relevant Project KPIs
				proposed in the baseline SOTA.		
	Progress, Results and Comments	Work for this component will start in project period 2.				
AVCC	Accuracy	MAE MSE	AudioCSRNet	DISCO	13.63 MAE	KPI-O2-E3-1 KPI-O2-E3-2 KPI-O2-E3-3
	Progress, Results and Comments	Achieved improvement of 4.28% over baseline.				
	Speed	FLOPS	AudioCSRNet	DISCO	30% Speedup while retaining 90% accuracy	
	Progress, Results and Comments	48% speedup retaining 93% of accuracy (14.58)				
ViAD	Accuracy	ROC Area Under ROC Curve		Street Scene and UCSD Pedestrian	62% frame AUC on Street Scene, 97.5% frame AUC on UCSD	KPI-O2-E2-1 KPI-O2-E3-1 KPI-O2-E3-2 KPI-O2-E3-3 iKPI-3-3
	Progress, Results and Comments	Performance achieved in UCSD: 99%				
	Speed	FLOPS		Street Scene and UCSD Pedestrian	30% Speedup while retaining 90% accuracy	
	Progress, Results and Comments	Reduction in number of FLOPs: 21% Reduction in model size: 71%				
VCC	Accuracy	MAE MSE	M-SFANet Transformer-based crowd counting	Shanghai Tech Parts A and B, World Expo 10 DISCO	11.00 MAE on average for sub-regions of the image	KPI-O2-E3-1 KPI-O2-E3-2 KPI-O2-E3-3
	Progress, Results and Comments	Achieved Performance in DISCO: 11.09 MAE				
	Speed	FLOPS	M-SFANet Transformer-based crowd counting	Shanghai Tech Parts A and B, World Expo 10 DISCO	30% speedup while retaining 90% accuracy	
	Progress, Results and Comments	Performance using EE at 3rd layer: 12.24 MAE (90.05% of performance retained)				
AAC	Accuracy	SPIDEr	Baseline of the AAC task	AudioCaps, Clotho, and data available in MARVEL	15% relative improvement on SPIDEr in DCASE setup (dataset, cross-validation setup)	KPI-O2-E2-1 KPI-O2-E3-1 KPI-O3-E4-1 iKPI-3-3 iKPI-4-1
	Progress, Results and Comments	Component was not developed for R1, and the work will start in project period 2.				
ASC	Accuracy	Classification accuracy (macro-averaged)	Baseline of the ASC task at DCASE Challenge	Publicly available datasets, e.g., dataset for ASC task in DCASE, and dataset available in MARVEL	10% relative improvement on accuracy in DCASE setup (dataset, cross-validation setup)	KPI-O2-E2-1 iKPI-3-2 iKPI-3-3

Asset	KPI	Metric	Baseline SOTA	Datasets / Benchmarks	Expected result	Relevant Project KPIs
	Progress, Results and Comments	This component was replaced with AT for R1.				
SED	Accuracy	Segment-based error rate F1-score (1-sec segments)	Baseline of the SED task at DCASE Challenge	Publicly available datasets, e.g., SED data used at the SED task of DCASE, and dataset available in MARVEL	10% relative improvement on metrics in DCASE setup (dataset, cross-validation setup)	KPI-O2-E2-1 KPI-O2-E3-1 iKPI-3-2
	Progress, Results and Comments	Achieved metrics on DCASE2017 SED setup: 37.8% F1-score and 0.88 error-rate (relative increase of 15% in F-score and 11% decrease in error rate). Evaluation is done with macro-averaged metrics to score under-represented classes.				
SELD	Accuracy	Location-dependent error rate and F1-score and classification-dependent localisation error and recall	Baseline of the SELD task at DCASE Challenge	Publicly available datasets, e.g., SELD data used at the SELD task of DCASE, and dataset available in MARVEL	10% relative improvement on metrics in DCASE setup (dataset, cross-validation setup)	KPI-O2-E2-1 KPI-O2-E3-1 iKPI-3-2
	Progress, Results and Comments	Component was not developed for R1, and the work will start in project period 2.				
AT	Accuracy	Mean average precision	Baseline of the AT task at DCASE Challenge	Publicly available datasets, e.g., data used at the AT task of DCASE, and dataset available in MARVEL	10% relative improvement on metrics in DCASE setup (dataset, cross-validation setup)	KPI-O2-E2-1 KPI-O2-E3-1 iKPI-3-2
	Progress, Results and Comments	Achieved mean average precision (mAP) of 0.79 on DCASE2018 audio tagging dataset (FSDKaggle2018 dataset) resulting in a relative increase of 16% against DCASE2019 audio tagging baseline system evaluated on the same dataset.				
CATFLOW	Performance	GRN Dataset	Experiments on MARVEL data	GRN Dataset	Accuracy per class	KPI-O2-E3-1
	Progress, Results and Comments	**Classification Accuracy:** mAp values per each class • pedestrian = 0.5517 • car = 0.7225 • bus = 0.8048 • light goods vehicle = 0.7454 • heavy goods vehicle = 0.7010 • motorcycle = 0.6593 • bicycle = 0.6011 • average mAp for each class = 0.6927 **Latency:** the CATFlow algorithm is able to work at a rate faster than real-time, thus no latency is observed.				
	Scalability	Computational resources needed	Use a set of benchmarks (designed within project MARVEL)	GRN Dataset	Number of instances given traffic volume supported by resources.	
	Progress, Results and Comments	**Scalability in terms of Computational resources:** • Linux OS PC compatible • 4 CPU cores				

Asset	KPI	Metric	Baseline SOTA	Datasets / Benchmarks	Expected result	Relevant Project KPIs
		• 8 GB RAM • 50 GB Storage • For processing: Processing RTSP Stream at 1920*1080P at 20fps with ~80% accuracy. The required input rate is between 10 and 30fps • With every new stream: 1-2 more CPU cores, 3-4GB RAM, no added storage • Also needs: NVIDIA GPU with Tensor cores (e.g.: RTX series, T4) [GPU compute capability needs to be addressed before use] A typical high-power workstation, similar to the GRN Edge PC would be able to sustain 2 instances of CATFlow.				
TAD	Accuracy	ROC Curve, Area Under ROC Curve	Not Available	Possible Publicly Available Datasets and MARVEL Datasets	>53% accuracy in detecting relevant anomalies	KPI-O2-E2-1 KPI-O2-E3-1 KPI-O2-E3-3
	Progress, Results and Comments	Precision per vehicle type for anomalous high-speed vehicles: • Car: 97% • Light Good Vehicle: 82% • Motorcycle: 99% • Bus: 100% • Heavy good vehicle: 100% Precision per vehicle type for anomalous low speed vehicles: • Car: 83% • Light good Vehicle: 75% • Motorcycle: 56% • Bus: 73% • Heavy good vehicle: 76% All tests were performed on datasets from one camera; Mgarr Camera. More work needs to be done to generalise the method and take into account more characteristics of the vehicles.				

8.5 Optimised E2F2C processing and deployment

Table 25: Asset specific KPIs for the "Optimised E2F2C processing and deployment" subsystem (6.9 in D1.2)

Asset	KPI	Metric	Baseline SOTA	Datasets / Benchmarks	Expected result	Relevant Project KPIs
GPURegex	Efficiency	Throughput Execution time	CPU-only implementation	Converted text-based MARVEL datasets (if available), otherwise publicly available alternatives	At least 10% processing speed-up	KPI-O1-E1-2
	Progress, Results and Comments	GPURegex was evaluated using several public datasets that include captions extracted from audio and video, and that could resemble the output of AAC, when available. TAU proposed three captioning datasets, namely Clotho, AudioCaps, and MACS. The evaluation process of GPURegex is fully presented in D4.2. In this deliverable, we briefly describe the results achieved. For the Clotho dataset, we use a pattern file that was constructed by the keywords that were extracted by the captions. Each one of these patterns is a				

Asset	KPI	Metric	Baseline SOTA	Datasets / Benchmarks	Expected result	Relevant Project KPIs
		fixed string, with a total of 6,623 patterns. The resulting automaton has 20,788 state transitions. The average GPURegex processing throughput is 3,282 Mbits/second (among 30 GPURegex executions using the same pattern file and input file that is 1.3MBytes), while the average processing execution time (latency) is 3,500 microseconds, whereas the CPU version of Aho-Corasick achieves processing in 12,000 microseconds. For the AudioCaps dataset, the pattern file was constructed by keywords that were extracted by the captions. Each one of these patterns is a fixed string, with a total of 7,414 patterns. The resulting automaton has 14,465 state transitions. The average GPURegex processing throughput is 3,200 Mbits/second (among 30 GPURegex executions using the same pattern file and input file that is 3.5MBytes), while the average processing execution time (latency) is 2,000 microseconds, whereas the CPU version of Aho-Corasick achieves processing in 28,000 microseconds. For the MACS dataset, the pattern file was constructed by the tags of the YAML file provided in the dataset. Each one of these patterns is a fixed string, with a total of 16 patterns. The average GPURegex processing throughput is 6,748 Mbits/second (among 30 GPURegex executions using the same pattern file and input file that is 933KBytes), while the average processing execution time (latency) is 626 microseconds, whereas the CPU version of Aho-Corasick achieves processing in 3,500 microseconds. In all experiments conducted, GPURegex achieves better performance (i.e., at least 10% processing speedup), when compared to the CPU-only implementation.				
DynHP	Model compression performance	Accuracy of the compressed models; Compression rate	Original and uncompressed model	MNIST; Fashion MNIST; CIFAR-10	Up to 70% compression rates, with up to 15% of relative accuracy drop w.r.t. the original model.	KPI-O3-E2-1 KPI-O3-E2-2
	Progress, Results and Comments	Up to 88% of compression rate with <1% of accuracy drop on MLP models. Up to 50% of compression rate with <3% of accuracy drop on VGG and WRN28-1.				
	Accuracy of real-time inference on the edge devices	Frequency of correct real-time inferences	N/A	N/A	Increase by 20% the number of accurate and real-time prediction performed within a fixed memory budget.	
	Progress, Results and Comments	Simulated performance in terms of reduced FLOPS: 36% (WRN28-1 on CIFAR-10) 46% (MLP on MNIST) 12% (MLP on Fashion MNIST)				
FedL	Personalised federated learning improvement (performance) at least 10%.	Accuracy	Standard non-personalised federated learning (LEAF, FedML)	FEMNIST, Federated AI dataset	Regarding model accuracy, we will devise novel personalised methodology which accounts for personalised data mixtures, and hence will be fine-tuned to the local data distribution.	KPI-O2-E1-1 iKPI-1-1 iKPI-3-1 iKPI-12-1 iKPI-12-2
	Progress, Results and Comments	On the LEAF benchmark, NUS strategy exhibits 30% savings in terms of amount of communicated data for negligible difference in accuracy (2%). Personalised clustering achieves improvement over SOTA approach on a synthetic dataset in terms of client cluster structure identification and model sparsity.				

Asset	KPI	Metric	Baseline SOTA	Datasets / Benchmarks	Expected result	Relevant Project KPIs
	Usability; Diverse tools for complex/federated/distributed systems handling extremely large volumes and streams of data.	Usability, Robustness	FedAvg (LEAF, FedML)	FEMNIST, Federated AI dataset	Three different methodologies will be devised and tested: personalised federated learning, novel FL protocol accounting for communication failures ("complex environment"), and split learning (SplitFed).	
	Progress, Results and Comments	Three different methodologies are proposed: personalised federated learning, novel FL protocol accounting for communication failures ("complex environment"), and split computing, communication efficient algorithms for social learning.				
	Accuracy ratio (>95%) of training with respect to full dataset training.	Accuracy	Training with access to full dataset (impractical or impossible due to privacy and other restriction – e.g., data transmission /data centralisation).	FEMNIST, Federated AI dataset	Training over distributed and/or private datasets while maintaining similar performance levels.	
	Progress, Results and Comments	Training over distributed and/or private datasets while maintaining similar performance levels.				
	Reduced training time by at least 10% compared to standard approach.	Training time/speed	Standard protocol oblivious to communication failures and client dropouts.	FEMNIST, Federated AI dataset	An adaptive FL protocol that accounts for communication and node availability statistics thus effectively reducing training time compared to a standard protocol.	
	Progress, Results and Comments	An adaptive FL protocol that accounts for communication and node availability statistics thus effectively reducing training time compared to a standard protocol.				
Marvdash	Usability	Reduced effort to specify component deployments User experience (UE) while interaction with the UI	Similar deployments without the usability benefits of the user-facing front end	N/A	Make it straightforward for domain experts to interact with resources in the E2F2C platform without having to understand lower-level tools and interfaces.	KPI-O3-E2-2 iKPI-1.1 iKPI-12.2

Asset	KPI	Metric	Baseline SOTA	Datasets / Benchmarks	Expected result	Relevant Project KPIs
	Progress, Results and Comments	The overall assessment was conducted through a survey in the context of D5.2, where the users rated their experience. Without MARVdash, users would be required to write deployment files, specify the individual placement of containers in the continuum, and apply these configurations through the command-line interfaces of each system component. The survey included questions regarding the main functionalities of the MARVdash, its advantages and disadvantages compared to other Kubernetes dashboards, and the whole user experience. The survey was divided into three sets of questions. The first set referred to MARVdash-specific functionalities such as starting a service, using a template, uploading an image, etc. The questions of the second set compared MARVdash with other Kubernetes dashboards. Finally, the third set included all the questions from the User Experience Questionnaire that helped us compare MARVdash with other products regarding the user experience. The last set of questions was further grouped into six scales: Attractiveness, Perspicuity, Efficiency, Dependability, Stimulation, and Novelty. For each of those, a number was produced as an output of the benchmark. According to the results of this process, participants rated MARVdash main functionalities with averages in the range of 4.75 (lowest) to 6.22 (highest). These values belong to the "Very good" category (one of them belongs to the "Excellent") of the corresponding qualitative assessment. Moreover, the user experience part of the assessment showed that MARVdash's means are above average compared with a large number of other products. Based on that result, MARVdash could be successful in the market.				

8.6 HPC infrastructure

The HPC infrastructure provided by PSNC consists of the HPC supercomputer (Eagle) and virtualised private Cloud (LabITaaS). While the most computation-intensive tasks, such as model training, are performed on the Eagle supercomputer, we use Cloud infrastructure as a base for deploying the MARVEL software stack. Both infrastructures are equipped with management and orchestration systems that allow for flexible allocation of computing resources and various class storage services.

Eagle and LabITaas are being monitored on physical (servers, switches, state of links, etc.) and logical levels. From MARVEL's perspective, more important is logical level monitoring on which all basic components are tested and checked for proper behaviour. Monitoring and checking against agreed SLA levels is done using Zabbix software.

The tested modules of Eagle are:

- Eagle: aggregate value that reflects the overall availability level of the HPC cluster. This aggregate is calculated assuming that if any of the underlying services is not working at a given moment, the entire service is considered as not available. From a technical perspective, it is not always true, but it reflects the "worst case" scenario.
- NIS server: reflects the state of the IDP provider. Unavailability of this service is unlikely to affect ongoing processing on the cluster, but it will prevent interactive work and the start of new calculations.
- Slurm: reflects the state of the resource manager. Unavailability of this service prevents new jobs from starting and may cause disruption in currently running calculations.

- Subnet manager: the component responsible for managing a high-speed cluster network. Unavailability of this service will prevent applications from accessing fast storage and, in the case of MPI-based parallel applications, will cause fatal errors in processing jobs. Currently, the cluster consists of several Infiniband fabrics, so there are failure zones: not all processing is affected by this kind of error.
- UI: the component responsible for monitoring traditional user interface available via ssh service. In the case of this component stare of several processes are monitored both from basic functionality (work or does not work) and from performance perspective.

The tested modules of LabITaaS are:

- LabITaaS: aggregates values for the cloud infrastructure calculated based on all components. This aggregate considers the whole cloud as non 100% functional if any of the components is not available.
- cinder: block storage service for all virtual machines and volumes used in the cloud environment. Unavailability of this service may affect running VM (virtual machines) as it may (but not have to) indicate that VM lost access to the underlying storage.
- horizon: web interface used for human interaction for creation or manipulation of virtualised resources. Unavailability of this service does not affect running VMs but prevents the possibility of manipulating the resources using the WWW interface and does not affect any already configured services.
- keystone: basic identity service used by OpenStack components. Unavailability mostly affects the ability of components to communicate with each other but does not imply malfunction of VMs.
- neutron: network virtualisation service. Unavailability affects the possibility of changing or creating networking components in the Cloud and may also mean some problems with accessing some or all virtual machines from the external networks.
- nova: service responsible for the creation and manipulation of the VMs. Unavailability prevents new VMs creation.
- S3 RADOS GW: access to object storage via S3 gateway. This represents the availability of the S3 storage. It does not directly affect VMs but may mean a lack of data access for services that use this storage.
- subnet manager – management service for Infiniband network. Because PSNC cloud relies on high speed, low-latency Infiniband network for storage and networking, malfunction of this service means severe problems with all other services and VMs running in the environment.

Below are the graphs containing measurements from both infrastructures for July 2022. One can see that there are several columns listed. The first column holds the component's name, followed by the current state. The next column shows unavailability levels both in graphical and numerical format. The last column presents the current and threshold values of the SLA for a given module. Some modules are aggregates of multiple lower-level metrics, but in most cases, these reflect the redundant nature of each service.

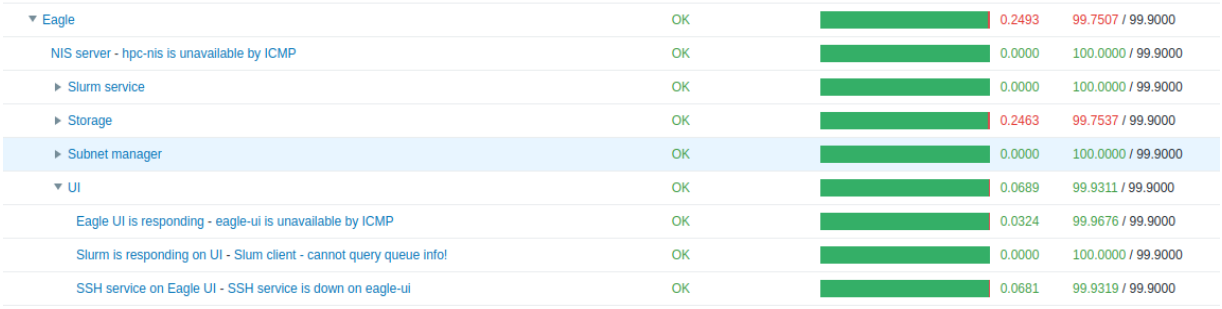

Figure 8: Eagle measurement results for July 2022

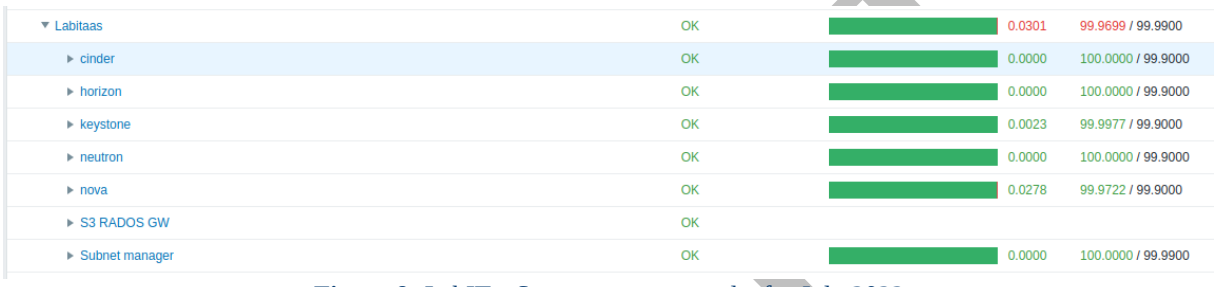

Figure 9: LabITaaS measurement results for July 2022

Table 26: Asset specific KPIs for the "HPC infrastructure" subsystem (6.10 in D1.2)

Asset	KPI	Metric	Baseline SOTA	Datasets / Benchmark	Expected result	Relevant Project KPIs
HPC Infrastructure	Reliability	Availability of the computing system [%]	-	Zabbix monitoring	90%	KPI-O3-E1-1 KPI-O5-E4-1
	Progress, Results and Comments	Figure 8 and Figure 9 show that Eagle and Cloud achieved much better results in this metric (99.75% and 99.97%, respectively) than expected.				
Management and orchestration of HPC resources	Reliability	Availability of the computing system [%]	-	Zabbix monitoring	90%	
	Progress, Results and Comments	Figure 8 and Figure 9 show that individual modules responsible for managing and orchestrating both infrastructures achieved much better results in this metric (> 99.9%) than expected.				

8.7 User interaction and decision-making toolkit

SmartViz is the core MARVEL component for decision-making and a main interaction point with the MARVEL user, concerning the real-time monitoring of a piloting environment (e.g., informed about ongoing events in the smart city). In addition, the user (both internal and external) can interact with the Data Corpus, where already captured and anonymised datasets are available. Furthermore, the user can visualise the datasets' details and download them.

Table 27: Asset specific KPIs for the "User interaction and d**ecision-making toolkit" subsystem (6.11 in D1.2)**

Asset	KPI	Metric	Baseline SOTA	Datasets / Benchmarks	Expected result	Relevant Project KPIs
SmartViz	Usability	Perception of: - time required: to accomplish a task - task completion quality Degree to which user's needs are satisfied	>80% positive perception and satisfaction. Results via 1-5 scaled questions	N/A (the decision-making toolkit will utilise data processed by the pilots	>85% positive perception. Results via 1-5 scaled questions	KPI-O3-E4-1 KPI-O1-E1-2
		Progress, Results and Comments	Wide-scale usability tests will be performed during our second Info Day (M24). Until then, the benchmarking measurements performed for the MVP (D5.2) apply here as well.			
	Scalability	Visualised data volume	Supporting browser friendly data volumes with no errors that prevent the reliability and performance of the interface		Supporting Big Data without errors that prevent the interface usability and performance	
		Progress, Results and Comments	The toolkit accepts a big volume of real-time and historical data streams per use case through the tailor-made adapters we developed and used. The adapters are used to utilise the available data, convert them into meaningful visualisations and extract valuable knowledge. In the current version the toolkit can support in each use case view in real-time mode 1000 items per minute, and up to 10,0000 items in batch mode. Therefore, R1 release of our tool ensures interface usability and nominal values of performance metrics.			
	Availability	Uptime	95%		No downtime during the operation of the tool	
		Progress, Results and Comments	There have been no reports of downtime by the SmartViz deployment in MARVdash by FORTH besides the development testbeds.			
	Reliability	Accuracy & Data Error Resilience	Interface responsive in case of data errors. Informative messages to users		No errors will prevent the interface from being responsive. No faulty information either	
		Progress, Results and Comments	The improvement, integration of new features and development of new visualisations of SmartViz is still in progress. During the next development phase of the toolkit the error handling mechanisms and the interpretation of errors via informative messages to the users, will be carried out in order to be evaluated.			
	Performance	Response time	Initial Page load: 9.92s (avg) Internal Page Loading: <1.25s (avg)		8-10 sec	
		Progress, Results and Comments	The average page load time is 6.78s. The performance will be re-evaluated upon the finalisation of the final version of SmartViz (M30).			

Asset	KPI	Metric	Baseline SOTA	Datasets / Benchmarks	Expected result	Relevant Project KPIs
MARVEL Data Corpus-as-a-Service	Scalability	Size of Data volume	N/A	N/A	Supporting scaling up the infrastructure without errors that prevent the corpus usability and performance	KPI-O1-E1-1 KPI-O5-E1-1 KPI-O5-E2-1 KPI-O5-E3-1 KPI-O5-E4-1 iKPI-11.1
		Progress, Results and Comments	The initial size of the Corpus was 100TB. Currently, at M22, the size of the Corpus is around 0.58PB. This will reach the 3.3PB by M36 as the project progresses. The current datasets include: - 10 piloting datasets - 6 open datasets - 420 augmented versions of the abovementioned datasets The required resources are continuously provided by PSNC, as more data are becoming available from the piloting environments.			
	Usability	Success rate (whether users can perform a query at all)	70%	N/A	80%	
		Progress, Results and Comments	Will be evaluated in R2 period			
		The time a query requires to be prepared by the user	8 to 10 sec	N/A	5 to 8 sec	
		Progress, Results and Comments	The users can interact with the Corpus via a web interface and filter/query information for datasets or their underlying snippets. This graphical interface is simple and easy-to-use.			
		Error rate	30%	N/A	20%	
		Progress, Results and Comments	Will be evaluated in R2 period			
		Overall user's subjective satisfaction (questionnaire)	>80% positive perception and satisfaction. Results via 1-4 scaled questions	N/A	>85% positive perception. Results via 1-4 scaled questions	
		Progress, Results and Comments	Currently, only the pilots and other MARVEL partners have used the Corpus. Their user experience is positive with several recommendations having considered during the web interface development. When the next version of the Corpus GUI will be ready, a dedicated workshop will be organised, involving external users and disclosing their opinion.			
	Performance	Time to query response	More than 10 sec	This factor can be affected by the total volume of data	5 to 10 sec	

Asset	KPI	Metric	Baseline SOTA	Datasets / Benchmarks	Expected result	Relevant Project KPIs
		Progress, Results and Comments	Will be evaluated in R2 period			
	GDPR-related criteria	SLAs satisfaction / violation	N/A	N/A	>90% of SLAs with no violations	
		Progress, Results and Comments	The related task T2.4 started at M20. As a baseline, an SLA-based continuous monitoring solution based on STS Security and Privacy Assurance Platform, is under development The platform will audit the operation of the system and assess wherever security and privacy aspects are violated or not. The initial considerations include scenarios of evaluating the access to Corpus resources by authorised users as well as the availability of the provided services.			

9 Project wide KPIs and Evaluation Parameters

This chapter deals with the evaluation of project-wide parameters and KPIs which are presented in three sections; Project scientific and technical objectives (section 9.1.1) and impact (section 9.1.2); Societal goals evaluation (Section 9.2) and Business goals evaluation (section 9.3).

9.1 Project KPIs

The current status of the project KPIs is reported here.

9.1.1 Project scientific and technical objectives

This section presents the main scientific and technical achievements of the project up to M22 toward the project objectives and is organised around the project objectives as defined in the DoA. For each of the objectives (5 in number), we report in a table the related KPIs, with pointers to the specific part of the DoA (WP, Task), detailed reports on the current achievements and plans until the end of the project.

9.1.1.1 Objective 1: Leverage innovative technologies for data acquisition, management and distribution to develop a privacy-aware engineering solution for revealing valuable and hidden societal knowledge in a smart city environment

Table 28: KPIs status update – Objective 1

KPI-O1-E1-1	Different kind of resources to be discoverable: ≥ 3	Achieved
This KPI is addressed within the scope of WP2 and tasks T2.1 and T2.3 where diverse IoT devices are currently connected to the framework. The KPI has been fulfilled as more than three types of resources i) microphones, ii) cameras, iii) drones, and iv) crowd-sourced mobile applications are used, while in some use cases weather and accident data will be considered as well. Related data has been stored in the Data Corpus as well. Next steps for the upcoming period include to keep investing if any other besides the "traditional" resources can be used under the scope of MARVEL.		
KPI-O1-E1-2	**Increase of data throughput and decrease of access latency by 10%**	**In progress**
Actions to increase throughput and decrease access latency were taken in the scope of WP4, T4.1. An audio processing board was customised for use with 4 IFAG MEMS (Micro Electro-Mechanical Systems). A custom firmware was written for it which increases the processing speed by x5, thereby reducing the latency significantly. Two AudioHub Nano versions, one with 2 microphones and one with 4 microphones, were released. Both come with a microcontroller that allows to connect the microphones directly via USB sending the data via I2S. This facilitates the connection to a processing edge node. Furthermore, work on an AudioHub Nano version with 8 microphones and a WiFi connection for the data was initiated. This 8 microphones board are expected to be finished by M22, aggregating all 8 audios signal in only one synchronised data channel, eliminating the need for several data channel and decreasing the access latency. This board will enable the audio data processing on the edge, implementing algorithms such as noise reduction or beamforming or sound source location. This board features a PSoC (Programmable System on a Chip) 64 microcontroller that allows for Machine Learning (ML) algorithms to run directly on that microcontroller.		
KPI-O1-E2-1	**Execution time of data management and distribution improved at least 15%**	**In progress**
Execution time is defined as the time needed between a data action request and the actual results delivered from this action. It can be broken down into different time segments, such as data moving time, data processing time, computation time, inter-node delays, and others. This KPI is being addressed within the scope of WP2 by moving data (or computation) where and when needed satisfying real or non-real-time application constraints, via rigorous modelling of the given networking environments, and through the extension and application of traditional (combinatorial) optimisation solutions tailored to realistic application settings. During the reporting period, we managed to achieve improved data access latency via minimising the Apache Kafka replication latency metric in MARVEL-tailored simulation-based performance evaluation, through the		

HDD component in WP2, T2.2. During the reporting period, we also managed to achieve improved data loss performance via similarly minimising the Apache Kafka unavailability metric in MARVEL-tailored simulation-based performance evaluation, through the HDD component in WP2, T2.2. Improvements are reported in detail in D2.2. In order to further address the KPI goal, the T2.2 components have undergone a rigorous integration process. During the next period, execution times will be able to be measured also in a more interconnected, systematic manner.

KPI-O1-E2-2	Increase the number of different modality data streams that can be handled by 30%	In progress

In order to increase the number of different modality streams handled by the MARVEL platform, we have incorporated the ability to handle large binary streams, such as audio-visual data. This was an ability that the components of the project lacked but it was also crucial for the project's success. To that end, we experimented with frameworks that offer multi-cloud storage and retrieval capabilities and can handle such data efficiently. The outcome of this process has been incorporated into StreamHandler and integrated into the project's Data Management Platform, under Task 2.2. Over the next period, we will monitor the audio-visual capabilities of the DMP in terms of performance and scalability.

KPI-O1-E2-3	Increase the speed of the fusion process by at least 20%	In progress

Up to now, we focused on increasing the speed of consumption of data streams by incorporating a module that sits between the Kafka message broker and the Elastic Search (ES) Repository of the DFB. The plan within T2.2 was to develop and integrate a mechanism so that streams can be stored into ES for later processing and queries in batches. An initial version of this mechanism has been developed, integrated and tested in the context of the initial integrated prototype of the MARVEL framework R1, as reported in D2.2 and D5.4. This is expected to vastly increase the fusion speed, which was previously based on a simple pooling mechanism. Furthermore, apart from supporting the process of fusing inference results from diverse sources at the ES repository, the mechanism between Kafka and ES also supports the fusion of multiple distinct inference results from a single origin into unique composite counterparts. The latter capability has been applied for a selected range of inference results produced by specific AI components (i.e., ViAD, AVAD). In these cases, the original inference results arriving at the DFB are detected anomalies that refer to specific instants in time, i.e., they have no extent in time. The fusion mechanism merges consecutive anomaly events into composite counterparts that refer to a period in time. This process is foreseen to occur during the inference results ingestion and the output of fused events is made immediately available at the Kafka message broker and at the Elastic Search repository. The real-time operation of this procedure provides immense improvements in terms of speed compared to a periodic, schedule-based post-processing operation. An extension of this procedure to apply to inference results produced by other AI components (e.g., SED, AT) is currently being examined to be implemented. Finally, the overall fusion speed is foreseen to be measured in the context of T5.4 activities.

KPI-O1-E2-4	Improve data distribution in relevant device resource usage at least 15%	In progress

Relevant device resource usage will consider selected application-oriented resources (such as energy consumption or processing time or memory allocation). This KPI is being addressed within the scope of WP2 by deriving decentralised approaches, designing low-complexity algorithms, and investigating resource consumption/solution accuracy level trade-offs. During the reporting period, the exact resources of critical importance to be optimised have been identified. The plan to fulfil this KPI is involving the design of smart ways of resource usage, based on the modelling limitations of the considered technologies. During the reporting period, we managed to achieve an improved number of special nodes/equipment via minimising the Apache Kafka used brokers metric in MARVEL-tailored simulation-based performance evaluation, through the HDD component in WP2, T2.2. The detailed results are presented in D2.2 This metric indirectly also reflects the energy consumption rate in an Apache Kafka cluster.

KPI-O1-E3-1	Number of incorporated safety mechanisms (e.g., for privacy, voice anonymisation) ≥ 3	Achieved

This KPI is addressed under T3.1, T4.2 and T4.3. In T3.1, FBK has been working on the development of audio and video anonymisation solutions. A first set of solutions has been released and YOLOv5 object detection + blurring is used for video anonymisation (for faces and plates). With regards to audio anonymisation, the current solution is based on shifting the LPC poles and is purely signal processing based which makes it suitable to work in real-time scenarios while being deployed on-premise. In T4.2, AUD has been working on improving its Voice Activity Detection (VAD) model. The model has been upgraded and re-

trained with respect to a novel state-of-the-art approach. As real-time inference on edge devices is one of the considered safety measures, the new model has been integrated into devAIce which is conceived to function in real-time on high-end edge devices in order to avoid transferring any sensitive audio data to the successive layers. During the first months of the project, AUD and FBK have been working on improving their newly developed technologies, and by the end of the integration period, the new VAD module has been integrated with AudioAnony, forming together the audio anonymisation pipeline. This composite component was deployed on edge devices (Raspberry PIs) in the use cases where audio data is present, UNS and MT, and a successful connection and synergy with the rest of the project components (DatAna, DFB, AudioTagging, etc) has been achieved. In T4.3, FORTH has developed, configured and integrated two security-enabling components: EdgeSec VPN, providing end-to-end security across all entities of the MARVEL framework, and EdgeSec TEE, providing security of edge and other devices by ensuring internal device integrity of the code and data.

To resume, by M18, the KPI is achieved and currently, 6 safety mechanisms are exploited: VAD, AudioAnony, VideoAnony, EdgeSec VPN, EdgeSec TEE and deployment on edge devices for audio components, as VideoAnony is too resource-hungry to be deployed on-premise.

As a future work, AUD and FBK will keep improving their anonymisation technologies, other optimisations like using GAN-based face-swapping for video anonymisation will be exploited, and benchmarks will be carried out when annotated pilot data will be available.

KPI-O1-E3-2	The end-to-end data flow from the edge to the cloud, will be 100% encrypted	In progress

After the end of the first year of the project, we had to divide the component "EdgeSec" into "EdgeSec VPN" and "EdgeSec TEE" with respect to the two underlying functionalities offered. EdgeSec VPN is implemented within the scope of Task 4.3, part of which offers the necessary security features to the E2F2C framework. EdgeSec VPN secures the data flow transmitted through MARVEL components. The component has been posted on the MARVEL image registry and can be downloaded by anyone interested in using it. EdgeSec VPN is gradually developing to cover all layers of MARVEL (Edge, Fog, Cloud) and every participating component and pilot. Currently, EdgeSec VPN secures the communication channel of all three pilots: UNS (Edge, Fog, Cloud), GRN (Edge, Fog, Cloud) and MT/FBK (Fog, Cloud). The KPI is still in progress and it will be achieved in the upcoming months when the communication channels between each and every component will be encrypted by EdgeSec VPN. The detailed status of the implementation of EdgeSec VPN is described in D4.2 (submitted in M18). The updates will be detailed in D4.5.

KPI-O1-E3-3	Video and voice anonymisation expected to improve by at least 10%	In progress

As the original description of the KPI did not specify what the 10% improvement refers to, we thus specified it with reasonable interpretation. Considering the different nature when processing video and audio data, we further break down the KPI for audio anonymisation as the 10% information increase in the audio content compared to complete speech removal, while for video anonymisation, we aim at a 10% reduction in computational complexity for video anonymisation as video processing is in general computationally demanding. This KPI is related to enabler E3 and is achieved under WP3 and WP4. Specifically, it will be implemented through T3.1 and T4.2, and in relation to T3.5 for the model compression.

Up to now, we have achieved baselines audio anonymisation using: speech removal via voice activity detection, signal processing voice modification, and voice conversion approaches. The latter is based on a sequence-2-sequence strategy and allows increasing the amount of speech information while preserving privacy, although robustness to noise and computational complexity are still open issues for deployment in the prototype.

Regarding video anonymisation, we have already achieved a baseline anonymisation solution with a blurring technique followed by faces/car plate detection. We have achieved a GAN-based baseline for anonymisation through face swapping, which has been evaluated on public face-related image datasets. Moreover, we have benchmarked state-of-the-art methods on our newly dataset which is built on top of public face-related image datasets but with the emphasis on varying face resolution, which is commonly observed in CCTV (Closed-Circuit Television) videos. Finally, we have reduced the complexity of a baseline GAN-based model via quantisation techniques, while being able to address the KPI, its performance still requires thorough evaluation in the next period.

For the next period, the goal is to finalise the development of the voice conversion method for deployment, so that audio anonymisation can achieve the targeted 10% increase of information in the anonymised audio

content. Then for the video anonymisation solution, we will further evaluate the compressed model and further investigate other techniques to compress the model in collaboration with related MARVEL partners (in T3.5) so that the complexity reduces by 10%, with the eventual goal of running the component at the edge.

9.1.1.2 Objective 2: Deliver AI-based multimodal perception and intelligence for audio-visual scene recognition, event detection and situational awareness in a smart city environment

Table 29: KPIs status update – Objective 2

KPI-O2-E1-1	**Standard non-personalised federated learning improvement (performance and speed) at least 10%**	**In progress**
This KPI relates to WP3, T3.2. Regarding training speed, an adaptive FL protocol was proposed and developed within T3.2 that accounts for communication and node availability statistics thus effectively increasing training speed compared to a standard protocol. For clients that exhibit heterogeneities, the protocol achieves a significant improvement on a synthetic dataset in terms of the training speed over the baseline FedAvg. On the LEAF benchmark (widely adopted and most commonly used benchmark for FL), specifically the MNIST dataset, FedL with NUS strategy reaches an accuracy of 88%, which is only 2.3% (=88/90) lower relative to the accuracy of 90% of the baseline algorithm FedAvg, while at the same time achieving significant savings in the number of communicated data. Specifically, in the same experiments, to reach the indicated accuracy, FedAvg requires transmission of 120Mbs in total, while FedL NUS requires in total less than 80Mb of transmitted model data, achieving savings of more than 33%. In the future, we plan to verify the improvement on further benchmark datasets. Regarding performance improvement, a novel methodology for client model personalisation based on convex clustering was developed in the first period of the project, which accounts for personalised data mixtures, and hence achieves tuning the models to the local data distribution to achieve improved model accuracy. We will continue to work on implementing and further testing this approach.		
KPI-O2-E2-1	**Average accuracy enhancement for audio-visual representations and models at least 20%**	**In progress**
AU developed a method for Audio-Visual Crowd Counting (AVCC) which provides 3.06% improvement (in mean absolute error) against the state-of-the-art. AU also developed an efficient method for Visual Crowd Counting (VCC) employing Early Exit Branches to reduce the parameter count of the resulting model. The resulting models achieved improvement up to 11.11% (in mean absolute error) compared to the competing baseline. Moreover, AU developed methodologies for improving the training of deep neural networks through better initialisation of their parameters and by increasing the diversity of representations learned by different neurons in each layer, which improved performance compared to the baselines on visual data classification. Detailed description of these methodologies is provided in D3.1.		
KPI-O2-E3-1	**Increase the average accuracy for audio-visual event detection by at least 10%**	**Achieved**
TAU developed a Sound Event Detection (SED) method based on a light-weight Convolutional Recurrent Neural Network (CRNN) obtained by replacing Convolutional Neural Network blocks with depth-wise separable convolutions, and replacing recurrent neural network blocks with dilated convolutions, and by utilising data augmentation techniques to diversify data. The resulting model led to a 56% increase (in macro-averaged F1-score) compared to the competing method. TAU developed the SED method based on audio tagging approach, where tagging is applied inside consecutive segments to get sound event detection output. The system was based on Convolutional Neural Network (CNN) architecture, and the resulting model led to a 76% increase (in macro-averaged F1-score) compared to the competing method. Detailed description of these methodologies is provided in D3.1.		
KPI-O2-E3-2	**Increase the average accuracy for unsupervised audio-visual event detection at the edge by at least 10%**	**In progress**
AU developed an Unsupervised Visual Anomaly Detection (ViAD) method based on structured parameter pruning on Memory-augmented Deep Autoencoder which improves performance by 1.2% while reducing the		

model size and its number of floating-point operations (FLOPs) by 71.78% and 21%, respectively. Detailed description of this methodology is provided in D3.1.

KPI-O2-E3-3	Decrease the time needed to identify an event by at least 30% of current time	In progress

FBK developed edge solutions for SED based on compact neural models with low parameter and MOPS counts. These solutions are suitable for low-resource pervasive devices. Being able to detect sound events very close to the microphones they eliminate the latency due to data transfer to the cloud. TAU developed a Sound Event Localisation and Detection (SELD) method based on self-attention on learned audio features. The resulting models have almost double the number of parameters compared to the competing method. However, they were benchmarked to be 2.5 times faster than the competition during the inference due to parallelisation achieved with the self-attention blocks, i.e., they lead to a decrease in the inference time of 40%. AU developed an efficient method for Visual Anomaly Detection (ViAD) which employs structured pruning for reducing the parameters of a Memory-Augmented Deep Autoencoder network. The resulting model led to a reduction of the parameter count equal to 71.76% and a decrease of the FLOPs equal to 21%, while slightly improving accuracy (AUC). AU and UNS developed a Dynamic Split Computing method which determines the optimal neural network split for achieving the fastest execution time considering the structure of the neural network, the data batch size, and the transmission channel data rate. Detailed description of these methodologies is provided in D3.1.

9.1.1.3 Objective 3: Break technological silos, converge very diverse and novel engineering paradigms and establish a distributed and secure Edge-to-Fog-to-Cloud (E2F2C) ubiquitous computing framework in the big data value chain

Table 30: KPIs status update – Objective 3

KPI-O3-E1-1	Number of novel algorithms and tools utilised from diverse multi-domain technological areas ≥ 3	Achieved

After the project's first period, we can already state that this KPI is achieved. The MARVEL project takes advantage of a significant number of novel algorithms and tools which allow for the adoption of the Computing Continuum paradigm with a fluid integration of resources at all layers of E2F2C and along the entire data path. Utilised algorithms and tools come from diverse multi-domain technological areas corresponding to the individual layers of MARVEL architecture.

Data Management and Distribution Subsystem

- BroMin and BroMax are the two heuristic algorithms that Hierarchical Data Distribution (HDD) currently offers for Apache Kafka topic partitioning. Published in the proceedings of IEEE CITS 2022 conference[13].
- DatAna tool repurposes the NiFi ecosystem to enable E2F2C data management in terms of communication, moving, and transforming.
- Data Fusion Bus (DFB) tool is a customisable product for data fusion for multiple modality data streams between several connected components and permanent storage. The tool is under development.
- PSNC, an IT technology partner provides data infrastructure within the cloud and services for HPC infrastructure. In addition, the raw data and project results are stored at the data repositories located at the Polish data centre (BST, Poznan, PL).

Optimised E2F2C Processing and Deployment Subsystem

- DynHP uses an incremental pruning method for convolutional neural networks.
- FedL uses a novel non-uniform sampling FL strategy, allowing for flaky updates from federated clients. This strategy is implemented and offered as a custom strategy component for the Python flower federated learning framework.

[13] T. P. Raptis, A. Passarella, "On Efficiently Partitioning a Topic in Apache Kafka", 2022 International Conference on Computer, Information, and Telecommunication Systems (CITS2022), Athens, Greece.

E2F2C infrastructure

- PSNC offers novel HPC infrastructure (classified in the TOP500 list) and virtualised private cloud with an OpenStack web interface tool allowing for a flexible allocation of computing resources and various class storage services connected with the HPC system.

Audio, Visual and Multimodal AI subsystems

- Text Anomaly Detection (TAD) is a novel anomaly detection tool under development.
- Visual Anomaly Detection (ViAD) is a novel object-based visual anomaly detection tool.
- Visual Crowd Counting (VCC) implements a novel method published in the proceedings of the BMVC 2021 conference[14].
- Audio-Visual Crowd Counting (AVCC) implements a novel method for integrating audio and visual features in early exits. Published in Neural Networks journal[15].

KPI-O3-E2-1	Model compression algorithms to achieve 70% compression rates, without a noticeable degradation of accuracy	In progress

The actions performed in T3.5 for achieving the KPI regards the research for improving the methodology on which the DynHP (Dynamic Hard Pruning) component is based. The methodology has been successfully tested on State-of-the-Art benchmarks architectures and datasets. Precisely, the current version of DynHP prunes several Neural Network Topologies (i.e., MLP, WideResNet, and VGG). Regarding performance, it achieves up to 88% of compression with a <1% of accuracy drop for MLP topology and up to 50% of compression with ~3% of accuracy drop for VGG and WideResNet28. The application of DynHP to more complex Network topologies like the one developed within MARVEL for Audio-Visual Crowd Counting and Visual Crowd Counting is under investigation. Next step will be about exploring alternative compression approaches suitable for some of the model types developed in MARVEL and, if possible, identifying lightweight models to be trained that are able to meet the KPI objective.

KPI-O3-E2-2	Optimise performance (prediction accuracy, time-to-decision) of DL deployment by 20%	In progress

In the context of T3.4 of WP3, FORTH has defined a software architecture spanning from E2F2C that enables distributed execution of DL tasks, and with efficient use of execution resources as a major concern. The implemented E2F2C framework (Kubernetes cluster along with MARVdash dashboard on top of it) enables DL task distributed execution, making use of Kubernetes mechanisms such as Taints, Tolerations, and Node Affinity.

The optimisation goal of this KPI will be achieved with the future functionality of MARVdash, planned to be implemented in the second half of the project lifetime. MARVdash will be able to match the task resource requirements to the various execution sites available in the MARVEL distributed environment. Being able to match task resource requirements to the various execution sites, we will enable improvements both in performance, particularly time-to-decision, as well as in the sophistication of the DL models being deployed, thus enhancing prediction accuracy.

KPI-O3-E2-3	Increase accuracy levels of real-time observations at the edge devices by 20%	In progress

In the context of T3.5, techniques for optimising ML algorithms for deployment at the edge have been developed. The strategy to reach this KPI includes (i) development, training and testing of ML models, (ii) model compression to fit the restricted resource profiles of edge devices, and (iii) rely on E2F2C environment for the deployment of ML libraries tuned for low-resource, edge computing devices - e.g., PyTorch/Caffe, Tensorflow/TensorflowLite, SciKit learn, LibSVM.

The difference between the needed FLOPs of the compressed and uncompressed ML models can indirectly increase the accuracy levels. The considered benchmarks showed a FLOP reduction of 36%.

KPI-O3-E3-1	Realise a secure computing framework at all the processing layers	In progress

[14] A. Bakhtiarnia, Q. Zhang, and A. Iosifidis, "Multi-exit vision transformer for dynamic inference", British Machine Vision Conference (BMVC), 2021.

[15] A. Bakhtiarnia, Q. Zhang, and A. Iosifidis. Single-layer vision transformers for more accurate early exits with less overhead, Neural Networks, 2022.

After the end of the first year of the project, we had to divide the component "EdgeSec" into "EdgeSec VPN" and "EdgeSec TEE" with respect to the two underlying functionalities offered. EdgeSec VPN is implemented within the scope of T4.3, part of which offers the necessary security features to the E2F2C framework. EdgeSec VPN secures the data flow transmitted through MARVEL components, whereas EdgeSec TEE can secure the execution of applications that process sensitive data. Currently, EdgeSec VPN is deployed and secures a large portion of the network communications between the three layers (Edge, Fog, Cloud) and the participating components. EdgeSec TEE's deployment is still in progress, due to its very specific technical requirements. The details of the development and deployment of the two components are detailed in D4.2.

KPI-O3-E4-1	Detailed insights to more than 5 hidden correlations	In progress
This KPI is linked with the design and development of the decision-making toolkit (started at M10) and the visualisations depicted there. It is naturally linked to Enabler 4 and a direct output of T4.4. However, T4.2, T3.3 as well as the whole WP6 where the real-life experiments are being performed are of key importance since they feed the visualisations (T3.3 and T4.2) and evaluate them (WP6) respectively. This KPI is being closely monitored since the MVP release, where end-users had the opportunity to experiment with the decision-making toolkit. Since then, the DMT was developed in a way to facilitate finding hidden correlations in MARVEL's data in the next releases. In the next release, we are considering in collaboration with the pilots, to use visualisations (i) to compare the results among the outputs of different tools and algorithms, (ii) to investigate the role of time and weather in the appearance of traffic anomalies as well as in the appearance of unusual events in the use cases related to public safety, (iii) to investigate what happens at the same period of time across different parts of the city based on historical data. More areas for investigation and use of visualisations in order to find hidden correlations will be identified as we use the tool and further develop it to address all the pilot use cases. Progress will be checked every three months, in order to be able to judge whether to insist on investigating an area or change direction and try a different hypothesis.		

9.1.1.4 Objective 4: Realise societal opportunities in a smart city environment by validating tools and techniques in real- world settings

Table 31: KPIs status update – Objective 4

KPI-O4-E1-1	More than 10 trial cases to showcase framework's capabilities	In progress
Within T1.3 of WP1, 10 use cases were identified, based on the pilots' functional and non-functional requirements each of them including multiple trial cases (see pilot descriptions). The use cases are described in detail in D1.2. 5 of these use cases have been finalised and deployed during the first period of the project (M1-M18). This KPI is related to enablers E1 and E3. The KPI is still in progress as the use cases will be implemented and executed under WP6. In the upcoming months, FBK will keep monitoring the KPI under T6.1. We expect to achieve the KPI with the final prototype release.		

KPI-O4-E2-1	Identify at least 20 dependent and independent verification and validation variables for the system	Achieved
The experimental protocol, described in D1.2 as a result of the efforts under T1.3, includes the definition of the experimental variables to be used to verify the progress of the project and validate the final framework. Tables in Section 6 of D1.2 list both the dependent variables related to the use case KPIs, as well as the independent KPIs used to assess the achievements of each MARVEL's asset. Table 6.12 summarises the experimental parameters, their variables and how they are measured. These have been defined starting from the pilot requirements as well as accounting for the expected asset innovation reported in the DoA, also in terms of Technology Readiness Level (TRL). The evaluation procedure will be updated in WP5 (benchmarks) and WP6.		

KPI-O4-E3-1	Execute the trial cases in at least 2 real life smart city environments	In progress
Trial cases are experiments implemented on the ground. These are combinations of the use cases defined in Section 1.3.4.5 of the DoA, following output from T1.3 and technology developed throughout the project. This KPI is related to WP6, T6.2. The activities started on M12 and are connected with several other Tasks and WPs (WP2-T2.1 and T2.2, WP3-T3.2, T3.3, T3.4, and T3.5, WP4-T4.1 and T4.2, WP5-T5.1, and WP6-T6.3). MT, GRN and UNS deployed 5 of the 10 use cases defined in D1.2 (MT1, MT3, GRN3, GRN4 and UNS1). The data collected had been provided and shared with the MARVEL consortium. MT, GRN, and UNS, to successfully fulfil this KPI, planned and realised the pilot-focused meetings with technical partners in order to share		

information and discuss the issues related to data acquisition (WP4), data storage (WP2, WP3 and WP4), analytics/AI/ML (WP3), infrastructure (WP5), integration (WP5), visualisation (WP2 and WP4), demonstrators (WP6) and evaluation (WP6). The MARVEL's experimental protocol (D1.2) and the MARVEL MVP (D5.1) were the starting point for the alignment of all data providers, in order to ensure smooth and adequate execution of the experiments. In the upcoming months, MT, GRN, and UNS are planning to deploy the other 5 use cases (MT2, MT4, GRN1, GRN2, and UNS2) and to improve the 5 use cases of the first reporting period (MT1, MT3, GRN3, GRN4, and UNS1).

9.1.1.5 Objective 5: Foster the European Data Economy vision and create new scientific and business opportunities by offering the MARVEL Data Corpus as a free service and contributing to BDVA standards

Table 32: KPIs status update – Objective 5

KPI-O5-E1-1	**More than 3.3PB of data made available through a Corpus-as-a-Service**	**In progress**
In the first phase of the project, the main components and their interconnections were developed. In order to achieve more than 3.3PB of data to be available through the Corpus, a continuous augmentation of the data that are ingested into the Corpus must be performed. Therefore, under T2.3 in WP2 an augmentation strategy is developed including SOTA techniques in terms of augmenting audio/video data. Furthermore, a series of scripts were developed in order to automatically inject to the Corpus data that are received from the data pilots. However, due to limited data availability in terms of volume and velocity that we are currently facing, a portion of data has been injected and augmented into the Corpus. A decent amount of data, i.e., 0.58PB, has been currently uploaded in the Corpus, which will be gradually increased during the second period of MARVEL until it reaches the designated threshold. It is expected, that MARVEL data providers to bypass the difficulties and achieve the target of this KPI until the end of the project. In order to achieve the final goal of sharing more than 3.3 PB of data, for the next period, actions must be taken under two pillars: a) in accordance with T2.3, data augmentation techniques to boost and elevate the augmentation rate, and b) by creating a streaming data injection pipeline, in order to provide to the MARVEL data providers a different pathway of sharing/placing their data directly to the Corpus. The latter will further boost the injection data rate to the Corpus repository.		
KPI-O5-E2-1	**Release SLAs and consider all the relevant aspects, namely accessibility, operability, managing streaming and network, legal considerations, security, privacy and technical concerns**	**In progress**
In order to facilitate the latter KPI reflected under T2.4 (not started yet) in WP2 a monitoring mechanism of Data Corpus must be deployed for continuous observation of the latter in terms of security, privacy, and technical aspects. The main elements will start after M20 and the kick-off of T2.4. Nevertheless, during the first period of the project, a stable approach of including in Data Corpus a SOTA Security Information and Event Management has been deployed (Elastic Search) as a set of tools and services offering a holistic view of the Corpus information security. For the next period, in order to fulfil this KPI, actions must be taken towards two directions: a) create and release under MARVEL an SLA that will consider aspects such as accessibility, operability, security, privacy, streaming, and network considerations and b) create and deploy a mechanism that will analyse events received from the monitoring procedure of the Corpus and alert for potential violations/satisfactions of the SLA.		
KPI-O5-E3-1	**More than 5 SMEs used the Corpus**	**In progress**
The actions performed up to now to address this KPI that is connected to WP2 (T2.3, T2.4) were focused mainly on creating a robust set of functionalities in order to make it possible for MARVEL to share its own Corpus in a secure environment where innovative applications can be built by third parties. The approach followed so far was to develop, using SOTA open-source technologies, a secure environment for delivering the Corpus as a public, obtained free of charge and as-a-service, making it possible for SMEs and start-ups to build on top of these data assets and create new business by exploring extreme-scale multimodal analytics. Concerning this KPI, progress will be recorded in the second half of the project, and probably during the third year of the project when the Corpus SLAs will have been deployed and the overall solution will be more mature in order to be offered to external users. Towards this aspect and for the next period, the strategy that will be followed is to ensure and enhance the stability of the Corpus in terms of security and reliability and disseminate the Corpus functionalities through a variety of social events and dissemination activities for attracting potential SMEs for its usage.		

KPI-O5-E4-1	Maintain the corpus for at least one year after the end of the project	In progress
This KPI is going to be addressed within the scope of WP2 and T2.4 by delivering at the end of the project the Data Corpus-as-a-Service, including all the necessary maintenance procedures. To do so, an investigation must be performed in the following period for locating the mandatory objectives of maintenance management of the Corpus such as the optimisation of the reliability of the equipment and infrastructure that will host the Corpus, improvements and adjustments in the operational safety of the Corpus. Nevertheless, PSNC, which provides the infrastructure that is used by the Corpus, is committed to support this KPI after the end of the project.		

9.1.2 Impact

One of the main goals of MARVEL apart from the scientific and technical advancements to meet the project objectives is also achieving the expected impacts. To do so, the MARVEL consortium relies on an impact maximisation strategy that is based on four fundamental axes:

- **Axis 1 - Mission-oriented approach:** cracking specific challenges in the Big Data Value chain addressing rising challenges of the European Data Economy (WP1-WP7).
- **Axis 2 – Openness:** open-access sharing of knowledge and cross-fertilisation with other relevant EU-funded programmes and communities for multimodal event detection (WP1-T1.1, T1.2, T1.3, T1.4, WP7- T7.2, T7.3).
- **Axis 3 – Sustainability**: invest in research and innovation to produce new knowledge and advance existing one, ensuring sustainable growth for the technological advancements (WP6-T6.3, WP7-T7.2, T7.3).
- **Axis 4 - Public engagement**: engage citizens and secure their support to promote breakthrough innovation (WP7-T7.2, T7.3).

The expected impact of the MARVEL project is related to (i) the work programme, (ii) the innovation capacity, competitiveness, and growth, and (iii) standards. For each impact category, well-defined KPIs have been identified by the MARVEL consortium aiming to monitor the progress within the course of the project and measure the respective achievements. Below, for each impact iKPI the status update, the activities up to now and the strategy towards the successful completion are reported.

9.1.2.1 Impact related to the work programme

Table 33: KPIs status update – Impact related to the work programme

iKPI-1.1	At least three (3) tools for complex/federated/distributed systems handling extremely large volumes and streams of data	Achieved
As a result of the work carried out in T3.2, UNS has delivered FedL component implementing the non-uniform client sampling strategy that is particularly well-suited for federated training with a large number of participating clients with arbitrary data modalities, including multimodal data streams. Second, within T3.2 a novel methodology for client clustering within FL with a matching algorithm is developed that finds hidden cluster structures in an arbitrarily large pool of clients. The relevance of this methodology is to facilitate model search for a newly arriving client by checking only a small number of model candidates, which are provided by the tool, and hence significantly reducing the complexity of the overall process when very large number of clients/datasets are participating in the training. Third, a tool for Dynamic split computing for distributing DL architectures to two consecutive tiers of the E2F2C architecture is developed within T3.3 (AU, UNS). The tool can operate on DL models of large sizes (large number of layers) and is suitable for data of large volumes (such as video). Finally, within T3.2 a tool is proposed for optimising data exchange protocols in fully distributed systems with complexity exhibited both in topology (e.g., generic topology) and inherent system randomness (link failures, node failures). For such systems, we develop a tool for adaptively optimising communication frequencies of nodes'		

interactions/data transmission and observe important communication savings can be achieved with respect to a time-constant strategy, with negligible performance deterioration.

iKPI-2.1	**Reduced time-to-market (TTM) by at least 15% for the development of analytics in the MARVEL's decision-making toolkit based on the E2F2C MARVEL architecture**	**In progress**

The development of MARVEL's decision-making toolkit is part of the T4.4 activities. At M18, D4.3 has been published. The document contains the initial version of the MARVEL decision-making toolkit. Specifically, it illustrates the design process, user requirements, and demonstration results from selected use cases. The DMT will be completed during the next phase of T4.4 until the final deliverable D4.6 which will be released at M30. During the first period of the project, we discussed about how to measure iKPI-2.1: this will be done via questionnaires. The latter will provide a quantitative estimate of the effort that has been saved due to the use of the E2F2C MARVEL architecture, based on their previous experiences in relevant/similar settings. The steps toward the achievement of the iKPI-2.1 will be the following: prepare the questionnaires and send them to the DMT developers; collect the responses and analyse them in both qualitative and quantitative terms; share the results with the consortium for discussion; distil the learned lessons and best practices.

iKPI-2.2	**At least three (3) different cyber threats avoided due to E2F2C**	**In progress**

This KPI is directly connected with the expectations and goals of T4.3 and should be addressed in that context. The goal is to provide a quantitative measurement of the final results of this task (T4.3) with regard to data integrity and confidentiality. On that note, we should first deploy our solution as part of a real use case and then perform the evaluation with respect to the objective of iKPI-2.2. During this upcoming period, we identified a list of cyber-attacks that could potentially occur and respectively avoided. Towards that end, we will intensify our efforts in the upcoming reporting period to be in a position to prove and/or demonstrate the applied protection mechanisms.

iKPI-3.1	**Accuracy ratio (>95%) of training of deep learning models with respect to full dataset training**	**In progress**

On the LEAF benchmark (widely adopted and most commonly used benchmark for FL), specifically MNIST dataset, FedL with NUS strategy reaches an accuracy of 88%, which is only 2.3% (=88/90) lower relative to the accuracy of 90% of the baseline algorithm FedAvg, while at the same time achieving significant savings in the number of communicated data. Specifically, in the same experiments, to reach the indicated accuracy, FedAvg requires transmission of 120Mbs in total, while FedL NUS requires in total less than 80Mb of transmitted model data, achieving savings of more than 33%. We consider here the full dataset training to be the method that adopts a bucket sampling across the user data, that is, we consider SGD where a mini-batch is formed by sampling data from all users. Compared with this benchmark, we approach its accuracy to within 95%, 90% versus 88% achieved by NUS. We plan to verify the improvement on further benchmark datasets.

iKPI-3.2	**At least 20% reduction in LOC (lines of code) required due to new deep learning models**	**In progress**

All developed software implements easy-to-use implementations and interfaces based on widely adopted ML and DL libraries to reduce the number of lines needed for model training and for deploying the resulting models.

iKPI-3.3	**At least three (3) approaches tested for ML training algorithms**	**Achieved**

FBK investigated the use of Knowledge Distillation for training edge neural models. TAU employed data augmentation techniques specAugment and mixup to improve the training of the Sound Event Detection (SED) models. TAU also proposed a continual learning methodology for training Automated Audio Captioning (AAC) models, an active learning method for unsupervised training of deep learning models, and a cross-modal contrastive learning method for improving performance in audio data analysis using other data modalities. AU employed curriculum learning and Copycat finetuning for training deep learning models with early exit brances, and unsupervised learning of Convolutional Autoencoders. AU also developed methods for improving the parameter initialisation of deep learning models based on Discriminant Learning, as well as training the deep learning model parameters by increasing their diversity. Detailed description of these methodologies is provided in D3.1.

iKPI-4.1	**At least 15% sound/video data throughput improvement benchmarked against inference in the cloud (standard industry-validated process)**	**In progress**

Based mainly on the timestamps of intermediate steps of the inference pipeline, and partially on the monitorisation of the system, T5.4 partners should propose test cases/benchmarks for the key identified components. A new round to identify the usage of potential existing benchmarks will be carried out.

T5.4 will carry out the benchmarking of the selected components using the project infrastructure. Benchmarking to be carried out after M18 (first iteration) and closer to M30 (second iteration).

iKPI-4.2	**At least 15% sound/video data access time delays reduction benchmarked against inference in the cloud (standard industry-validated process)**	**In progress**

Based on the monitorisation of the time consumed in the inference pipeline via the usage of timestamps. Similar to the previous KPI, but focusing on time and data transfer reduction for real-time inference.

T5.4 will carry out the benchmarking of the selected technologies using the project infrastructure. Benchmarking to be carried out after M18 (first iteration) and closer to M30 (second iteration).

iKPI-5.1	**At least two (2) different trial cases for experimentation**	**Achieved**

The data providers GRN, UNS and MT have planned ten use cases spanning societal challenges in the areas of transport, personal safety and security, and crowd monitoring. Some of these trial cases can be potentially executed in the real-world over real-world data, and the rest in experimental settings with the recruitment of volunteers. GRN and MT had talks with authorities/private transport consultants and/or law enforcement agencies and /or legal bodies to ensure that the use cases address societal challenges and satisfy legal constraints.

We have three different trial cases (pilots) covering 10 different use cases of which 5 have been implemented for the R1 integration. Two use cases are from GRN, namely GRN Use Case 3 Traffic Conditions and Anomalous Events and GRN Use Case 4 Junction Traffic Trajectory Collection. Two use cases are from Trento. These are MT Use Case 1 Monitoring of crowded areas and MT Use Case 3 Monitoring of parking places. The last use case implemented is in UNS, which is the UNS Use Case 1 Drone Experiment.

WP6 is responsible for ensuring the implementation of the trial cases in Trento and Malta. Following the MVP implementation at M12, and the release of the MARVEL 1st integrated framework, T6.1, T6.2 and T6.3 continuously monitors the implementation of the use cases and in particular; follows closely the AI models and technology developed in the other WPs, and adjusts where necessary to make sure that the trial cases can be implemented; make sure any missing required regulatory permits are obtained to execute the trial cases; follow any shortage in hardware required in implementation; fine-tune the use cases according to feedback from stakeholders and other parties.

iKPI-5.2	**At least six (6) stakeholders engaged by the end of the project to further adopt MARVEL extreme-scale analytics tools and broader technology ecosystem**	**In progress**

Considering that in order to achieve this KPI, our targeted market segments should first become aware of the MARVEL extreme-scale analytics tools and broader technology ecosystem. Dissemination activities are very important for successfully accomplishing this KPI. Within Y1, the consortium partners have approached various stakeholders in plenty of events (for more details see D7.3) outlining the project objectives. Following up on the MVP, the first MARVEL Info Day was organised and it was attended by 96 attendees with several attendees coming from Municipalities, most of which expressed an interest to be kept in touch with the project consortium for further updates on progress. Other than the Info Day, the consortium partners showcased the MVP and disseminated MARVEL in more events (see details on our website), where they engaged with multiple stakeholders for adopting MARVEL. We have initiated discussions with 2 more smart cities in Greece (Trikala and Athens), we are part of the IEEE International Smart Cities Conference 2022 with a full-day workshop where we are presenting among others the MARVEL prototype in order to engage smart cities in adopting it and we initiated discussions with the Urban Technology Alliance, discussing the implementation of MARVEL in the testbeds of the smart cities that are part of them.

9.1.2.2 Impact on innovation capacity, competitiveness and growth

Table 34: KPIs status update – Impact on innovation capacity, competitiveness and growth

iKPI-10.1	**Integration of benchmarks, e.g., HiBench and SparkBench for Big Data; LEAF benchmark for federated learning for AI and machine learning; the DeFog benchmark for edge-fog.**	**In progress**
During the first year, an extensive study of existing benchmarks has been provided in the scope of D1.2 and the benchmarking strategy outlined in T1.3 and T5.4. There is no progress related to the integration or usage of these benchmarks so far, as the focus is the assessment of very specific components ready for the MVP and the methodology to measure system-wide metrics. The actual benchmarking will happen after the R1 and final integrated releases. Based on the monitorisation of the benchmarks, T5.4 partners will cross-check the potential usage of the benchmarks cited, as well as the ones mentioned in D1.2 (or emerging ones), especially for component benchmarking rather than for the system-wide benchmarks. A new round to identify the usage of potential existing benchmarks will be carried out after M18.		
iKPI-11.1	**Addressing at least seven aspects for efficient sharing of heterogeneous data pools; (i) accessibility, (ii) operability, (iii) streaming and network management, (iv) legal considerations, (v) security, (vi) privacy and (vii) technical concerns.**	**In progress**
For addressing the accessibility aspect, it has been developed, configured, and tested a flexible distributed framework for hosting heterogeneous data while providing access to it via respective interfaces (REST APIs), under WP2 and T2.3. The developed Data Corpus will be provided as a Service to external research and industrial communities. For addressing the operability aspect, it has been developed, configured, and tested a monitoring system to observe the overall performance of the latter in terms of usage resources (CPU, network, file disk etc.), under WP2 and T2.3. This monitoring system will be expanded in the second half of the project, so it can also be used for addressing aspects regarding legal obligations, privacy, and security. Also, the volume of the available data will be increased significantly, fostering the European vision toward smart cities and data economy.		
iKPI-12.1	**Reduced training time by at least 10% compared to standard approach**	**In progress**
An adaptive FL protocol was proposed and developed within T3.2 that accounts for communication and node availability statistics thus effectively reducing training time compared to a standard protocol. For clients that exhibit heterogeneities, the protocol achieves a significant improvement on a synthetic dataset in terms of the training speed over the baseline FedAvg. To complete this KPI, we will validate the approach on a benchmark dataset (LEAF).		
iKPI-12.2	**Increased performance in terms of response time, throughput and reliability compared to standard approach**	**In progress**
This KPI is under the lead of T3.2 but by its nature, it requires a joint effort between several different tasks of the project, most notably the tasks dealing with component deployment (T3.4), integration (T5.3) and benchmarking (T5.4). The plan is to carry out experiments based on R1 release with FedL and other relevant components, and use the methodology for monitoring relevant parameters (e.g., throughput, latency) defined within T5.4 and described in D5.4, to compare the response time and reliability of the FedL protocol and possibly other AI components with respect to a baseline approach.		
iKPI-13.1	**At least 3 industrial companies boosted exploiting MARVEL framework capabilities**	**In progress**
The iKPI-13.1 will be addressed via continuous monitoring of means to exploit MARVEL. Through continuous monitoring, we will be able to track the KPI and report on its level of completion by the end of the project.		
iKPI-13.2	**At least 3 smart cities directly engaged to MARVEL by the end of the project**	**In progress**
This is a continuous iterative process that aims to establish a network of cities that are up to date with the project's insights and results, ensuring a higher exposure and a sustainability plan. In the first 18 months of the project, consortium partners established advanced communication channels with multiple European smart cities regarding MARVEL. We are thus confident that this KPI will be met by the end of the project.		

iKPI-14.1	>8 SMEs providing or using data and/or data-related services (e.g. smart services) engaged by the end of the project	In progress
Considering that in order to achieve this KPI, SMEs (other than the ones included in the consortium) should first become aware of the MARVEL data and/or data-related services, dissemination activities are very important for successfully accomplishing this KPI. Within Y1, the consortium partners have approached SMEs in targeted events (for more details see D7.3) outlining the project objectives. Following up on the MVP the first MARVEL Info Day was organised and it was attended by 96 attendees coming also from the industry, professionals from innovation hubs, and a local network of enterprise representatives interested in leveraging AI solutions for multimodal big data analytics in their businesses, most of which expressed an interest to be kept in touch with the project consortium for further updates on progress. Other than the Info Day, the consortium partners showcased the MVP and disseminated MARVEL in more events (see details on our website), where they engaged with multiple stakeholders for exploring the MARVEL capabilities and even enhance their portfolios with the services developed in the project. Given that the MARVEL Data Corpus is key for the acceleration of progress regarding this KPI, we expect more progress in the second half of the project.		
iKPI-14.2	**>15% market share increment in MARVEL consortium technology providers**	**In progress**
To monitor the project's impact on the market share, ITML has included specific questions for the technology providers in order to be able to track the effect of MARVEL on the market share increment. This will be repeated in the course of the project and report on the level of completion of the KPI by the end of the project.		
iKPI-15.1	**>5% increment of penetration of IoT applications in MEMS microphone market**	**In progress**
In Y1, IFAG contributed to an increase in penetration of IoT applications in the MEMS microphone market by developing different boards that allow the connection of 2 to 4 MEMS microphones to any processing device very easily via USB, sending the data through the I2S protocol. In the next year, IFAG will continue to work on boards to facilitate the use of MEMS microphones. Currently, a board with 8 microphones which will send the data via Wi-Fi is under development. It will come with a powerful microcontroller that can run ML applications. Reports about MEMS microphones show that the current trend is that more and more revenue has been made with MEMS microphones in the recent past. This trend is expected to continue in the future. The highest increase in revenue was seen in the consumer electronic sector, with an expected increase from 168M to 264M USD in the years from 2020 to 2025. Translated to microphone volume, in this period is expected to grow from 870M to 1.9BN units. When looking into IoT smart infrastructures market, the expected growth is from 3.8M to 6.5M units in the 2020-2025 period.		

9.1.2.3 Impact on standards

Table 35: KPIs status update – Impact on standards

iKPI-16.1	At least four (4) formal specifications of standards and benchmarks that will be used and influenced	In progress
As in the case of iKPI-10, this KPI will be assessed in later stages of the project. T1.3 revised the SOTA on benchmarking related to the key technologies of the project candidate to produce solutions beyond the SOTA. T5.4 is also revising the proposed benchmarks, monitoring their evolution and considering potential new benchmarks that might appear. The different components of the MARVEL framework are using different benchmarks for assessing their functionality (reported in D5.2). More than 4 benchmarks or shared datasets for assessing performance have been used for that purpose so far. We will continue assessing this and potentially use some of the benchmarks listed in the DoA if necessary.		

9.2 Societal goals evaluation

Table 36 lists the societal goals that have been identified by the GRN and MT pilots and are reported in D1.2 (Table 7.2). It should be noted that most of these goals are difficult to evaluate during the lifetime of this project for two reasons: (a) the pilot implementations are limited in

space and time and any evaluation would be limited to such implementations; (b) a longer time period beyond the end of the project together with widespread deployment is necessary to evaluate whether the framework addresses some of the societal goals in the longer term. For these reasons, most of the evaluations will rely on surveying the perceived impact of the project on these goals. Table 36 summarises the evaluations which the pilot owners managed to collect till M22.

Table 36: Summary of the societal goals addressed by the pilots and the related measurement strategies. * indicates metrics whose evaluation would require a time span longer than the project

<table>
<tr><th>Pilot</th><th>Societal Objectives</th><th>Measurements</th></tr>
<tr><td rowspan="5">GRN</td><td>Increased uptake of active modes of travel, such as walking and cycling</td><td>Perceived changes in habits through targeted surveys
Mobility data*</td></tr>
<tr><td colspan="2">Progress, Results and Comments
An online survey to analyse to what extent the tools developed address societal challenges (increase safety and efficiency on the road, uptake of active modes) was mailed to potential respondents. The respondents had access only to a recorded demo of the tools in use and could not try it out in a live session. Four people answered the survey and two of them have experience in the transport industry. Below is a summary of the perceived impact per use case.
<u>GRN3: Traffic Conditions and Anomalous Events</u>
• The timely detection of anomalous events does not have a significant impact on active modes uptake
• Suggested additional factors that would encourage uptake of active modes of travel, such as walking and cycling.
o Law enforcement and better driving behaviour [To be partly dealt with in R2 period].
o Detection of congestion is useful if adequate infrastructure is provided.
<u>GRN4: Junction Traffic and Trajectory Collection</u>
• General agreement that long-term collection of data regarding traffic conditions and trajectories can help encourage the uptake of active modes of travel, such as walking and cycling.
• Suggested additional ideas on how this data can increase the uptake of active modes of travel:
o Instead of encouraging active modes of travel, it can help in the planning/designing for such modes.
o Origin-destination data [data not collected by the system]
o Times for walking vs driving can be calculated. If driving takes longer maybe better to walk. [data not collected by the system]</td></tr>
<tr><td>Increase of safety for vulnerable road users</td><td>Perceived safety through targeted surveys</td></tr>
<tr><td colspan="2">Progress, Results and Comments
Summary of the perceived impact per use case
<u>GRN3: Traffic Conditions and Anomalous Events</u>
• Strong agreement that Detecting irresponsible driving and obstructions on the road (detecting anomalous events) would increase safety for VRUs.
• On the other hand, traffic condition detection would have no impact on increasing the safety for vulnerable road users.
• Additional suggested anomalies to increase safety of vulnerable road users.
o Detecting jay walking, (crossing streets outside zebra crossings).
o Detecting vehicles on cycling paths [planned for R2].
<u>GRN4: Junction Traffic and Trajectory Collection</u>
• Strong agreement that the long-term collection of data regarding traffic conditions and trajectories can help increase the safety for vulnerable road users (i.e., cyclists and pedestrians).
• Additional comments on how the data collected can help improve the safety of vulnerable road users (i.e., cyclists and pedestrians).
o Trajectories especially for pedestrians and cyclists are useful to determine desired paths or critical crossing points.</td></tr>
</table>

<table>
<tr><th>Pilot</th><th>Societal Objectives</th><th>Measurements</th></tr>
<tr><td></td><td colspan="2">

Speed data is important as it will highlight over speeding issues.
Data may be used to discover which parts of the infrastructure are friendly to active modes of travel.

</td></tr>
<tr><td></td><td>Better education of road users.</td><td>Perceived usefulness through a survey</td></tr>
<tr><td></td><td colspan="2">
Progress, Results and Comments
Summary of the perceived impact per use case
<u>GRN3: Traffic Conditions and Anomalous Events</u>

Agreement that the detection of anomalous events will help in the education of road users.
Additional suggestions of how this system can be used in the education of road users.

Increase awareness of bad driving habits.
New and natural examples of the anomaly and how to go about it.
such a system should be linked to VMS (variable message signs) which are spread across the country (Malta) but do not offer any real-time information.

<u>GRN4: Junction Traffic and Trajectory Collection</u>

An overall weak agreement that long-term collection of data regarding traffic conditions and trajectories can help in the education of road users.
Additional suggested data that can be collected to educate road users.

Speed data needs to be collected and then drivers are alerted of their excessive driving speeds since speeding is a key contributor to collisions and injuries.

</td></tr>
<tr><td></td><td>More efficient use of the physical transport resources</td><td>Perceived usefulness through a survey</td></tr>
<tr><td></td><td colspan="2">
Progress, Results and Comments
Summary of the perceived impact per use case
<u>GRN3: Traffic Conditions and Anomalous Events</u>

Agreement that detection of anomalous events such as road accidents, obstructions on the road and irresponsible driving could result in more efficient use of the physical transport resources.
On the other hand, the timely detection of general traffic conditions [such as light or heavy traffic] would have little to no impact on the efficient use of the physical transport resources.
Suggestions on how traffic conditions detection can be used to help road users:

It can help by giving drivers the option to choose another route to their destination, rather than staying on the same road.

<u>GRN4: Junction Traffic and Trajectory Collection</u>

Agreement that long-term collection of data regarding traffic conditions and trajectories can help increase the efficient use of the physical transport resources.
Suggestions on desired additional data or statistics:

Data on traffic volumes, peak hours, AADT, [Annual Average Daily Traffic]
Data to help predict driving behaviours

</td></tr>
<tr><td></td><td>Better maintenance of the physical transport resources, for example, estimation of the volume of road traffic informs the kind and quantity of material to be used in road building.</td><td>Perceived usefulness through a survey
Mobility*</td></tr>
<tr><td></td><td colspan="2">
Progress, Results and Comments
Summary of the perceived impact per use case
<u>GRN3: Traffic Conditions and Anomalous Events</u>

Detecting traffic conditions or anomalous events would have little or no impact on the maintenance of the physical transport resources.

<u>GRN4: Junction Traffic and Trajectory Collection</u>

The data collected would have little impact on the maintenance of the physical transport resources.
Additional suggestions on how the data can help in the maintenance of resources.

</td></tr>
</table>

Pilot	Societal Objectives	Measurements
	○ The detection of a collision could automatically send a message to the road maintenance department to visit the site and fix the damage, once it has been reviewed by court experts or road safety auditors.	
MT	Better and quicker detection of potentially dangerous situations	Surveys for Local Police Number of citizens reports* Public opinion monitoring (via Mentionlytics[16])
	Progress, Results and Comments The surveys and data collection will be carried out or analysed during the R2 period.	
	Safer mobility in critical areas	Data from transport authorities* Surveys for Local Police
	Progress, Results and Comments The surveys and data collection will be carried out or analysed during the R2 period.	
	Better life quality and increased security for citizens	Annual reports such as the quality of life report of the newspaper "Il Sole 24 Ore"*
	Progress, Results and Comments The reports will be sampled during the R2 period.	

9.3 Business goals evaluation

The business goals and evaluation methods of said goals have been mapped and reported in D1.2 (Table 37) during the earlier months of the project. The business KPIs are related to a number of functional and non-functional KPIs, namely low latency and detection time, increased efficiency in data-drive planning/decision-making, end-user experience, citizens (or target groups) satisfaction, scalability. The business goals will be applied to the GRN and MT use cases, as shown in Table 37. So far, following the MVP and 1st integrated release (R1) these goals and their evaluation approach are still valid. However, since the system is not in production, goals such as revenue and profit increase, cost reduction etc. cannot be currently determined. The approach to evaluating these goals, and the goals themselves, will be revisited on M24 of the project during the scheduled benchmarking workshop, in order to remain up-to-date and current to the MARVEL framework development. The results will be reported in the final evaluation report, D6.4.

Table 37: Mapping of the business-related use case requirements into DataBench business metrics. The table includes actual pilots' requirements as well as high-level expected goals (Table 7.1 in D1.2)

Metric	GRN		MT	
	Requirement/Impact	Evaluation method	Requirement/Impact	Evaluation method
Revenues increase	GRN may improve the value of the current product making it more attractive for other potential clients.	Market Study	N/A	N/A
Profit increase	For the related public administration, more efficient use of taxpayer	System usability System efficacy Comparison w.r.t. the current situation	For the related public administration, more efficient use of taxpayer	System usability System efficacy

[16] https://www.mentionlytics.com

	funds towards safe and functional roads.		funds towards safe and functional public spaces.	Comparison w.r.t. the current situation
Cost reduction	Either cost reduction or an increase in productivity using data-driven decisions. Instantly available data can reduce the required human efforts (or allow better use of human efforts)	System usability System efficacy	Cost reduction due to increase in productivity and in more effective data-driven decisions. The use of the multiple sensors deployed in the city leads to reductions in the cost of human monitoring as well as in the reduction of social costs due to misbehaviour. Similarly, better use of the recorded data could reduce the planning cost for mobility improvement.	System usability System efficacy
Time efficiency	Reduced time in processing data for data-driven decisions. Quicker intervention for real-time monitoring scenarios.	End-user experience	Reduced time in processing data for data-driven decisions. Quicker intervention for real-time monitoring scenarios.	End-user experience
Product Service quality	Satisfaction of cyclists and other vulnerable road users. Improvement of data-driven design. Improvement of the internal end-user experience and administrator opinion on the use of the real-time sensor feed.	Surveys	Improvement of the internal end-user experience and administrator opinion on the use of the real-time sensor feeds. Improvement of the efficacy and efficiency of the data-driven decision-making process. Improvement of public opinion and reduction in citizens' reports.	Surveys to the Local Police Citizens' report and independent analyses Traffic-related measures Annual reports such as the quality of life report of the newspaper "Il Sole 24 Ore"
Business model innovation	More efficient transport networks that encourage green and active modes of transport, contributing to more responsible public spending and a better quality of life.	-	N/A	N/A

9.4 MARVEL Framework KPIs:

In addition to use case specific KPIs and asset specific KPIs, D1.2 (Table 6.12) defined a set of KPIs to enable the wholesome evaluation of the MARVEL final framework taking into consideration a variety of aspects, namely, technical framework performance, scalability and reliability of the framework, trustworthiness of the framework and usability and societal impact.

Currently, the focus is on the development and evaluation of the individual components towards the realisation of all the defined use cases. The evaluation of the overall MARVEL framework with respect to the KPIs defined in Table 38 will take place during the final year of the project. The rationale for this strategy is that (a) the evaluation of the MARVEL framework is highly dependent on the performance of the individual components and (b) a fully integrated version of the MARVEL framework is required to be able to evaluate it. Thus, we need first to (a) achieve the desired results on the component level and (b) achieve a fully integrated version of the framework (R2 or final integrated version of the MARVEL framework), so that we can then proceed with the evaluation of the MARVEL framework. For completeness, the table is included in this document but the results will be reported in D6.4 (due M36).

Table 38: MARVEL's framework evaluation parameters and variables (6.12 in D1.2)

Parameter	Variable	How to measure	Most Relevant subsystem
Performance	Accuracy	Objects/event detection accuracy	Sensing and Perception Audio Visual and Multimodal AI
		Pattern recognition accuracy	Sensing and Perception Audio Visual and Multimodal AI
	Real-Time processing	Latency	Sensing and Perception Data Management Optimised E2F2C processing
		Inference time/Response time	Sensing and Perception Audio Visual and Multimodal AI Optimised E2F2C processing
	Data Management	Data throughput	Data Management Optimised E2F2C processing
		Data loss rate	Data Management Optimised E2F2C processing
		Volume of processed data	Optimised E2F2C processing
Usability	Efficiency	End-User experience	User Interaction and Decision-Making
		Collected data	Sensing and Perception Data Management
		Latency	Sensing and Perception Multimodal AI Data Management Optimised E2F2C processing
		Reduced in domain material	Optimised E2F2C processing Audio Visual Multimodal AI
		Reduced training time	Optimised E2F2C processing Audio Visual Multimodal AI
	Efficacy	Number of reports to public authorities	Sensing and Perception
		satisfaction of external subjects (citizens, cyclists)	Audio Visual Multimodal AI
		Improved behaviour of target subjects (drivers)	User Interaction and Decision-Making
		Effective Planning	Sensing and Perception
Scalability	Data	Required Data Volume	Sensing and Perception

			Audio Visual Multimodal AI Optimised E2F2C processing
		Required bandwidth	Sensing and Perception Audio Visual Multimodal AI Optimised E2F2C processing
	Software	Overall memory footprint	Audio Visual Multimodal AI Optimised E2F2C processing
		Overall FLOP/MOPS	Audio Visual Multimodal AI Optimised E2F2C processing
		Modularity	Audio Visual Multimodal AI
		Latency	Sensing and Perception Data Management Optimised E2F2C processing
		Reduced training material	Audio Visual Multimodal AI Optimised E2F2C processing
	Hardware	Energy consumption	Sensing and Perception Optimised E2F2C processing
		Energy balance	Optimised E2F2C processing
		Hardware speed up	Optimised E2F2C processing
		Cost to add/maintain new sensors	Sensing and Perception
		Overall cost	ALL
		Hardware life time	Optimised E2F2C processing
Trustworthiness	Cybersecurity	Threats avoided	Security Privacy and data protection
		Amount of encrypted data in transit	Security Privacy and data protection
		Data security	Security Privacy and data protection
	Ethical, legal, and privacy-preserving requirements	Continuous evaluation of the data processing activities	ALL
		Respect of the secondary use of personal data	
		Respect of the limitations of research activities in non-EU countries	
		Informed consent for data processing activities	
		Prevent potential algorithmic bias	
		Prevent data misuse	
		Avoidance of the risk of mass surveillance	
		Implementation of anonymisation and pseudonymisation techniques	
	Responsible & Trustworthy AI	Human agency and oversight	Sensing and Perception Security Privacy and data protection Audio Visual Multimodal AI Optimised E2F2C processing Audio Visual Multimodal AI
		Technical robustness and safety	
		Privacy and data governance	
		Transparency	
		Diversity, non-discrimination and fairness	
		Societal and environmental wellbeing	
		Accountability	

Reliability (Robustness and Availability)	Hardware	Packet loss	Sensing and Perception Data Management
		Bandwidth limitation	Sensing and Perception Optimised E2F2C processing
	Operational conditions	Fault tolerance	Data Management User interaction and decision-making
		Environmental conditions	Sensing and perception Audio Visual and Multimodal AI
	Up time	Hoovering time	Sensing and perception
		Required visibility/environmental conditions	Data Management
	Data	Limited data access	Sensing and perception Data Management Audio Visual and Multimodal AI

10 Feedback from External Advisory Boards and External Stakeholders

The methodology that MARVEL is following to deliver a solid scientific prototype for E2F2C ubiquitous computing for increased intelligence in a smart city environment extends the typical cycle (requirements, development, and evaluation) to a more agile approach that relies on four phases: (i) the *Baseline phase* (M1-M8), (ii) the *Innovation phase* (M9-M18), (iii) the *Experimentation phase* (M19-M30), and (iv) the *Consolidation phase* (M31-M36). The Baseline phase was the driving force of the project where the MARVEL architecture and the experimental protocol were defined, and a SOTA update and an initial market analysis were conducted. This phase provided direct input to the subsequent phases. The Innovation phase started at M9 and included the investigation of different approaches for providing the project's individual components, the delivery of the MVP, which was planned early in the project timeplan and acted as a proof-of-concept demonstrator, and finally, the 1st Integrated prototype of the MARVEL framework at M18. The Experimentation phase started at M19 and the goal is to converge MARVEL technological innovations and execute field trials of the developed use cases while continuing the development of the remaining use cases that will be ready with the 2nd release of the MARVEL framework planned at M30. Finally, during the Consolidation phase that is expected to start at M31, the goal is to release the MARVEL Data Corpus and fine-tune MARVEL's framework.

In all the above phases, the feedback from external to the consortium partners plays a fundamental role in the design of the MARVEL framework. The external parties are grouped into two categories, i.e., MARVEL External Advisory Boards (EAB) and external stakeholders. Regarding the first category, MARVEL has established the MARVEL Advisory Board (AB) and the MARVEL Ethics Advisory Board (EB) with external independent experts aiming to receive valuable feedback on the project achievements and to bring excellence. In the second category, MARVEL is relying on the feedback received from external stakeholders to design a cutting-edge framework that will be useful for smart cities that aim to improve the citizens' quality of life. External stakeholders can be public authorities, Traffic Managers, Law Enforcement Agencies, data scientists, engineers, architects, technical project managers, citizens and other people who are focused on smart city technologies, products and services.

According to the MARVEL timeplan, the feedback is more or less collected after every release of the MARVEL framework in various events that the partners are organising and is taken into consideration during the design of the next release. In more detail, during the Innovation phase and with the release of the MARVEL MVP at M12, the first-year meetings with the AB and EB and the MARVEL MVP Info Day took place. The 2nd Info Day and the second-year meetings with the EABs are planned to be organised during the Experimentation phase and more specifically after the release of the 1st Integrated framework and by the end of M24 (December 2022) and the third-year meetings with the EABs and the 3rd Info Day during the Consolidation phase, after M31 and until the end of the project at M36.

In the following sections, the feedback that has been collected up to now, the actions that MARVEL has undertaken to uptake this feedback and the future plans for the events that are coming next are presented.

10.1 MARVEL External Advisory and Ethics Boards

The main task of the MARVEL Advisory Board (AB) is to provide external, independent analysis and recommendations on the project achievements and to bring additional competencies towards a full achievement of the MARVEL objectives. The responsibilities and

duties of the AB include connecting the project outcomes with potential users of the developed solutions, other projects and research initiatives, policy-makers, and standardisation bodies, following the project development and providing necessary feedback, and contributing significantly with fresh ideas regarding the challenges and opportunities from the emerging research and from an industrial perspective. The MARVEL Advisory Board consists of four (4) independent members external to the MARVEL consortium:

Table 39: External Advisory Board members

Member	Position	Company
Yannis Theodoridis	Professor	University of Piraeus, Greece
Mr. Darko Pekar	CEO	Alfanum doo Novi Sad, Serbia.
Mr. Antoine van Ruymbeke	President	MILSET Europe
Mrs. Nuria de Lama (joined at M17)	Consulting Director	IDC4EU

Apart from the Advisory Board, MARVEL has also established the MARVEL Ethics Board. The main task of the EB members is to oversee, advise, assess and, when applicable, raise concerns to the PC and consortium partners on relevant ethical issues within the project, with a special focus on the processing of personal data (PD). The MARVEL EB is composed by three (3) independent members external to the MARVEL consortium:

Table 40: External Ethics Board members

Member	Position	Company
Dusko Martic	Head of Content Protection and DPO	United Media Group
Anastasiya Kiseleva	Data Protection Expert & Legal Researcher	Vrije Universiteit Brussels
Julia Pshenichnaya (left at M17)	Cognitive process automation practice lead	IBM France
Dr. Ghasan Bhatti (joined at M18)	Director & IoT-NGIN Project Coordinator	Capgemini

The MARVEL EB is assisted by three consortium members:

- **Mr Djordje Djokic** (PN), MARVEL's Data Protection Manager (DPM);
- **Prof. Dragana Bajovic** (UNS), MARVEL's Scientific & Technical Project Manager (STPM);
- **Ms Despina Kopanaki** (FORTH), member of the MARVEL's coordination team and WP9 Leader.

During the first year, MARVEL organised two special events with its advisory boards:

On the 28th of April, 2021 the first MARVEL EB meeting was successfully held among the MARVEL consortium and the 3 members of the EB. The purpose of the meeting was to receive the EB's feedback and guidance on relevant privacy, security and ethical issues within the project. The EB members, after being informed of the project objectives and technical approach,

the smart cities test cases as well as the ethics challenges to address, provided their feedback, advice and guidance mainly on the secondary use of PD, the research activities in the non-EU country and the anonymisation techniques. Table 41 below summarises the feedback received by the EB members during the EB KOM and the actions that the MARVEL consortium is undertaking to address these comments and suggestions.

Table 41: MARVEL EB Feedback

Name of the EB member	Question/Comment/Suggestion	How MARVEL will work towards addressing the comment/suggestion
Dusko Martic	Decision-making process – what are the roles of people involved in the process, is the output fed back to the system, is response time the goal?	MARVEL develops methodologies for supporting both real-time and medium to long-term decision-making. With respect to real-time, the type of decision-making is an 'alert' type – raising an alarm and similar and which will only be considered as a supporting tool for human operators in the command centre who, with proper authorisations in the respective organisations, perform the actual decision-making. MARVEL defined specific KPIs related to the need to decrease the response time. Real-time response time is one of MARVEL targets.
Dusko Martic	Open data will be offered to the whole community or to specific sectors?	One of the main goals of MARVEL is to release Data Corpus-as-a-Service aiming to drive research and innovation in audio-visual data processing and analytics and will empower Smart City authorities to achieve more secure and responsible societies and to foster EU Data Economy. All data are being properly anonymised and a legal framework is developed to consider all the ethical and privacy concerns. MARVEL Data Corpus is developed within T2.2, shared within T2.3 and legal, ethics and privacy compliance are being continuously guaranteed within T2.5 and WP9.
Anastasiya Kiseleva	What will be the resolution of the images coming from the cameras? Will this affect the anonymisation that will be applied?	It depends on the use case. Within MARVEL a variety of resolutions with the cameras that are or will be deployed by the data pilots are offered. Datasets have been examined and appropriate anonymisation techniques are applied to guarantee that no sensitive information is exposed. The anonymisation techniques are being developed within T3.1 and T4.2. Data pilots also applied standard market anonymisation techniques before releasing data for training. Detailed report on the anonymisation techniques reported in D9.2, and D4.1 and will be provided in D3.3 (M24), D4.4 (M24).
Anastasiya Kiseleva	What will be the kind of events that MARVEL is going to monitor? Is it continuous monitoring or case by case?	The goal of MARVEL is to support continuous monitoring; however, it depends on the actual implementation. Extensive list of the events has been defined by the use case providers. The purpose of processing this data is not to identify a specific individual but to clarify and categorise what is the nature of events. Anonymisation techniques are being used and are further developed within the project-focused tasks.
Anastasiya Kiseleva	Merging audio and video data – Which kind of data will be merged? Raw data, anonymised data or both?	The process that is followed is to independently anonymise audio and video data. Merging is taking place after the anonymisation step. In order for MARVEL to achieve its goals, data might need to be processed separately since in

		some cases events are captured only by microphones or only by cameras.
Dusko Martic	Is there a technical way that you can exclude a range of human voice in audio data? Exclude all data of human voice from all the sounds gathered by the microphones?	SOTA solutions exist that can remove human speech up to some extent depending on the environment and other conditions. MARVEL considers these techniques as an initial step for anonymisation of audio data. However, within MARVEL further research is being conducted under T3.1 and T4.2 to develop techniques to de-identify the speech. Moreover, T3.1 develops methods and algorithms capable to run at the edge and ensure protection of PD in general audio-visual recordings. There are several events that can only be captured via audio (e.g., siren, ambulance, people screaming). Currently, the collaboration with FBK and AUD with AudioAnony and VAD components is able to detect speech segments and either completely remove or anonymise them.
Dusko Martic	Misuse of technological outcomes by malevolent actors. Was there any consideration of misuse by commercial entities with unintended consequences?	Two different outcomes: Research data and technologies that are being developed. The consortium is carefully examining the data that are becoming open to minimise potential misuse and also the development of privacy-preserving technologies. The EB is monitoring and advising the consortium activities to ensure that privacy-enhanced techniques are in place and also the data that will be released will not violate the privacy of the data subjects.
Anastasiya Kiseleva	Proposal of AI regulation. It might be important to take the requirements into consideration in advance. Suggestion: explore the requirements and define what kind of issues the technologies could have.	Within the scope of T1.3 and D1.2, a detailed analysis of the ethics guidelines for Trustworthy AI was performed. The guidelines were translated into requirements and integrated into the MARVEL framework evaluation KPIs. Within MARVEL, we will perform an exercise to check how these guidelines can be adopted and followed, reporting in parallel the results of the respective KPIs. T2.5 paves the way for the responsible AI approach set out in T1.3, while WP5 will perform a continuous alignment with the responsible AI planning and guidelines set out in T1.3.

On the 16th of December, 2021, the first MARVEL Advisory Board meeting was successfully conducted. The meeting was held online. Representatives from all consortium members were present along with the three Advisory Board members. The purpose of the meeting was to introduce the MARVEL project to the AB members and present the main achievements up to now aiming to receive valuable feedback for the next steps. The meeting consisted of four parts: (i) MARVEL project vision, objectives, and outcome, (ii) MARVEL technical solution and innovations, (iii) MARVEL smart city test cases, and (iv) MARVEL MVP. The AB members through a fruitful discussion provided their feedback and recommendations aiming to ensure high quality and excellence in the project. Their comments were focused on the challenges that MARVEL is facing related to data availability, integration of 35 components, complete realisation of the E2F2C infrastructure, privacy preservation, and ethics compliance. Table 42 below includes several focused points and comments that the AB members raised regarding specific aspects of the framework, together with accompanying reply summaries that were provided and will be further updated and enriched during the course of the project.

Table 42: MARVEL AB Feedback

Name of the AB member	Question/Comment/Suggestion	Reply summaries during the AB meeting from the consortium
Antoine van Ruymbeke	Is it possible to perform real-time operations at the edge?	Edge processing is one of the key project objectives, specifically real-time operation of AI inference. To enable real-time inference at the edge, we are working on several parallel fronts, such as ensuring availability of the appropriate edge hardware (e.g., NVIDIA Jetson Nano for DL, or RPis for light-weight models), model compression (e.g., SED@Edge, DynHP) and distributed NN architectures with early exits at the edge. Specifically, regarding the early exits, there is a concrete methodology – called dynamic graphs, to decide how deep one can go in the topology of the network to provide the response, which we will be working on as well.
Antoine van Ruymbeke	Monitoring of data throughout end-to-end?	The GA sets initial approaches/methodologies on how to monitor and benchmark end-to-end throughput by placing monitoring points at certain locations in the complete E2F2C workflow. This initial approach is to be elaborated and enriched within T5.4.
Yannis Theodoridis	Are there any specific targets in terms of data availability?	Yes, the final release of MARVEL Data Corpus aims to provide 3.3 PB of AV data.
Antoine van Ruymbeke	Have you considered the possibility of a feedback loop in the workflow?	In the context of the Drone experiment use case specifically (the workflow of which was presented and discussed at the AB meeting), we have considered exploiting the received inference results to make informed decisions on the drone trajectory (e.g., move the drone closer to the events of interest). More generally, feedback loop in the workflow will be instrumental to achieve optimised online deployment of tasks and services in the complete E2F2C framework based on various performance tracking indicators: e.g., inference accuracy of specific deployments, response time, etc.
Yannis Theodoridis	What is the expected improvement for the GRN pilot specifically?	Currently, only cameras are used to monitor and analyse traffic conditions. By using additional modality – audio, we expect to increase the accuracy of detecting certain events (e.g., distinguish between bicycles and motorcycles), which is of special importance in the presence of occlusions or low visibility conditions. The idea is to cross-check both systems to ensure higher system reliability/accuracy. Also, currently, the data is being used only a posteriori and we want to extend the system analytics to real-time operation as well.
Yannis Theodoridis	Regarding deployment of the analytics – Edge, Fog or Cloud, there are in general two options: to decide beforehand and to decide in real-time. Are you considering real-time deployment as well?	Yes. The idea of the optimised E2F2C processing and deployment subsystem is to use various performance tracking indicators: e.g., inference accuracy of specific deployments, response time, including hardware availability, to decide online the placement locations of AI tasks, including possible reallocations/redeployments, newly acquired or trained AI models as the model database is enriched, etc.
Antoine van Ruymbeke	Are you considering the possibility to add new components at the front-end of the platform?	As envisioned, the framework provides a full processing pipeline: from the data sources to the user interactions - visualisations and decision-making toolkit. We are using

		APIs for each of the internal components to ensure possible additions of new components.

On September 1st and 2nd, 2022, Ms. Nuria de Lama, member of the MARVEL AB since M17, was invited by the PC to attend the review preparation meeting that took place in Heraklion, Crete. The meeting was organised two weeks before the official Interim Review meeting of the project. The consortium together with Ms. de Lama had the opportunity to work in an interactive way on all the review material, i.e., presentations of all the sessions and demonstrators of the pilot use cases and the individual components. Ms. de Lama with her extensive experience in RIA projects provided useful insights and valuable feedback during the two days meeting and significantly helped the partners to improve the material for the review. Valuable feedback on the exploitation and long-term sustainability of the MARVEL framework for the second reporting period was received.

As indicated in the GA, one meeting per year for each of the EAB (i.e., AB and EB) will be scheduled aiming to update the members on the progress and the main achievements of the project. The comments will be collected after the end of the meetings and the actions of the consortium to address them will be reported in the yearly progress reports (D8.4, D8.5) and in D6.4 (Final assessment report and impact analysis) that is due M36.

Note that Table 41 and Table 42 are continuously updated to include the concrete actions of the consortium with respect to the comments received by the AB members. The goal of these tables is to better monitor the progress of the comments/feedback received by the EB and AB members.

10.2 Feedback from external stakeholders

10.2.1 MARVEL Minimum Viable Product Info Day

The 1st MARVEL MVP Info Day was successfully held on January 28th, 2022, virtually hosted by FBK. There were 96 attendees coming from Municipalities, and the industry, as well as academic experts in the field of ICT solutions for smart cities, professionals from innovation hubs, and a local network of enterprise representatives interested in leveraging AI solutions for multimodal big data analytics in their businesses. More specifically 60.47% of the participants were from research and academia, 7% from PAs, 9.3% were smart city service providers, 14% were project partners or relevant stakeholders in H2020, and the rest equally divided between governmental agencies, Big Data/IoT/AI-relevant association, and general public. During the event, several related poll questions were launched to collect the feeling of the audience about the technologies presented and feedback on the project motivation, strong and weak points, and possible future directions.

The objectives of the MARVEL MVP Info Day were:

- Raise awareness about the H2020 EU-funded MARVEL project.
- Present the MARVEL Minimum Viable Product related to the Malta smart city test cases.
- Collect feedback from smart city decision-makers, i.e., Public Administrations (PAs), law enforcement agencies, and industry specialists, on the MARVEL framework and the proposed technical approach. This feedback will contribute to shaping future steps within the project.
- Enlarge the community with an active interest in the project.

- Explore opportunities for continuous collaboration between the project and the relevant stakeholder (industries, PAs, governmental agencies such as law enforcement, traffic managers, etc.) in the area.

The event started with a welcome session by the Director of the Digital Society Centre of FBK, Dr. Marco Pistore, who presented how the strategy of FBK is in line with the aim pursued by MARVEL. Mr. Giacomo Fioroni, head of the Smart City project of the Municipality of Trento followed with a presentation tackling the question "Why is MARVEL relevant to the city of Trento?" as an example of why the idea behind MARVEL can be applied in other smart cities in Europe. FORTH as the PC of MARVEL, provided an overview of the project, highlighting the motivation, vision and objectives of MARVEL. MARVEL Scientific & Technical Coordinator (UNS), followed with the scientific and technical presentation of MARVEL highlighting the MARVEL architecture and the main pillars of the project.

After this introductory part of the project, the workshop focused more on the experimental real-life use cases from the two smart cities. MT and FBK presented the defined test cases for the smart city of Trento while GRN presented the respective test cases for the smart city of Malta. Finally, the demonstration of the MARVEL MVP, which was the main topic of the day, took place facilitated by MARVEL Integration Manager (ITML) and ZELUS responsible for the visualisation of the MARVEL framework. The workshop concluded with a brainstorming and discussion session.

During the Info Day, the participants had the opportunity to exchange ideas and provide feedback, by answering a series of poll questions. The questions spanned from investigating the audience's interest in MARVEL technologies and their application for the citizen quality of life improvement, which aspects were considered most innovative or relevant for smart cities, up to the identification of the main barriers or risks perceived and preventing their use in the attendees' organisations, and in general of the aspects recognised as most challenging for the project.

The project received very encouraging feedback regarding its potential to improve citizens' well-being in a smart city (>82% of positive answers). This also emerged later from the replies received from the polls during the presentation of the pilot test cases, i.e., 76% of the participants considered that the pilot in Trento can improve citizens' well-being, security, and safety; similarly, 83% evaluated that the roads test case, addressed in Malta, will help with data-driven decisions and consequently improve citizens' life.

The audience also provided feedback on the most innovative aspects identified in the architecture. The preferences were equally distributed between the audio-visual AI components with synchronous processing of audio-visual streams, the E2F2C continuum processing and optimised deployment, the audio and video anonymisation at the edge and the MARVEL Data Corpus-as-a-Service. Similarly, we asked which of these technologies were perceived as most relevant for big data applications in smart cities with a similar balance. There is anyway a slight prevalence towards AI technologies for processing multimodal streams and their anonymisation at the edge (see results of Q5 and Q6 below).

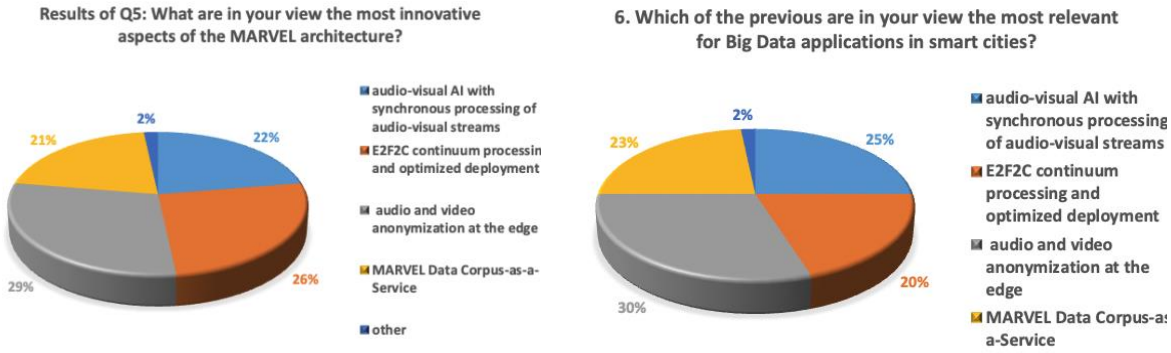

Figure 10: Poll questions during the 1st Info Day regarding the most innovative aspects of MARVEL and the most relevant for Big Data applications in smart cities

The audience was also invited to express their views on the challenges that MARVEL may face. In this open question, many respondents mentioned privacy as the main challenge but also similar concepts such as anonymisation, safety, and security. Other challenges identified were data availability, big data management in real-time, data labelling, bias in data, integration of components and operations in the framework, process optimisation, analytics but also energy-efficiency, functionalities, and privacy in the different layers.

Privacy is a clear aspect that emerged in several questions and poll answers, i.e., the concern for citizens' privacy and ethics, although MARVEL addresses this point very seriously both from the technology and legal point of view. Privacy is considered a challenge for several reasons; organisations may be reluctant to share data, as well as it may be very difficult to convince authorities or citizens that their privacy is respected. The consequence is often that major challenges arise in offering open data to the scientific and industrial community to grow their knowledge on several urban phenomena and related technology and business.

Additional ideas for both the cities of Trento and Malta were requested from the audience. The responses revealed that for Trento, many participants highlighted the need for smart parking and public transport management, while for Malta the focus was more on technologies for traffic management and awareness/monitoring.

The importance of having information available about activity on roads and open spaces in day-to-day life was considered "extremely important" by approximately 45% of attendees, while 41% considered it "somewhat important" and the rest remained neutral. Nevertheless, services related to green mobility (car/bike-sharing, public transport improvements), car parking and traffic are perceived as important in a smart city. Moreover, several suggestions for additional smart city services that can potentially be addressed in MARVEL were related to sustainable mobility and mobility in general.

The aim to present MVP at the Info Day was to receive feedback at a very early stage of the development cycle of the MARVEL framework to detect shortcomings and obstacles and to have sufficient time to correct the path. The use case selected for the MVP was the GRN use case "Junction Traffic and Trajectory Collection", which focuses on the analysis of long-term data to extract road users' behaviour (both vehicles and pedestrians) and also gather traffic statistics in road network junctions monitored by street-level cameras. This use case exploits a combination of MARVEL components and functionalities (e.g., vehicle identification and tracking, sound event and crowd counting.

During the MVP demonstration, it was explained how these technologies could be used by smart cities for decision-making regarding traffic, urban roads, etc. This is achieved through

the availability of relevant statistics followed by navigation and filtering options, as well as details about the presence of vehicles and pedestrians and their behaviour in a specific road junction of interest. The information can be downloaded or displayed with easy-to-understand plots and graphs.

Similarly, to the previous sessions, poll questions were launched and the feedback collected was very encouraging. Participants felt that the tool presented was easy to use and learn as indicated in the carts in Figure 11.

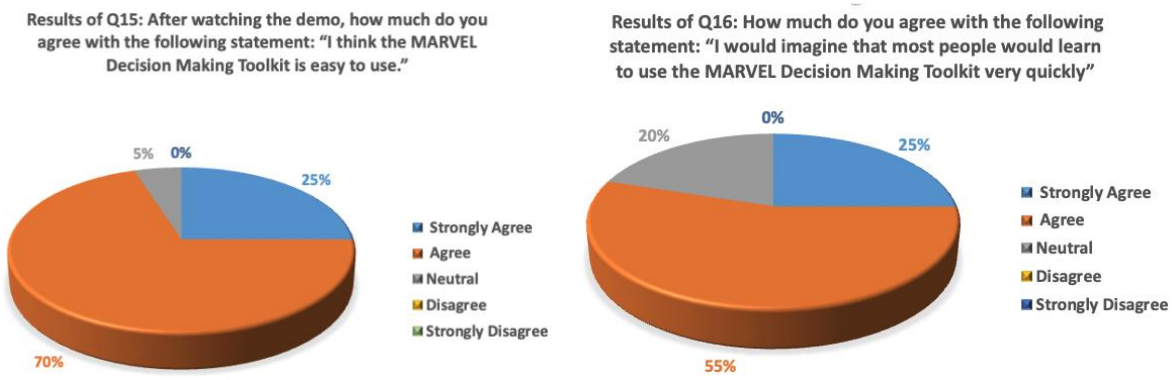

Figure 11: Poll questions during the 1st Info Day regarding the MARVEL Decision-Making Toolkit

Through an open question, we also collected some desired additional functionalities to consider for the next releases of the MARVEL framework such as crowdsourcing, the inclusion of more sensors, a forum for interaction, automated processing, real-data flows, real-time analytics and alerts, federated learning, analysis of preferred path based on the category of the vehicle, etc. Finally, the vast majority of the participants, i.e., 83%, felt that MARVEL will actually manage to improve citizen's well-being in a smart city as depicted in Figure 12.

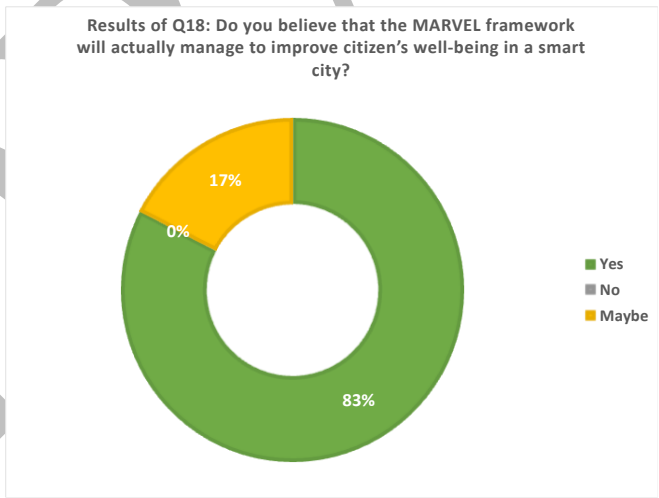

Figure 12: Poll question during the 1st Info Day – Will MARVEL framework actually manage to improve citizen's well-being in a smart city?

Overall, the feedback collected was very useful for the MARVEL consortium and helped us to better shape the next steps and achieve the next release by integrating the suggestions received during this Info Day. The conclusion was that there are two main directions that we need to pay attention for the next releases, i.e., the concerns of participants about violating citizens privacy and ethics compliance issues, and technical challenges/recommendations for the overall MARVEL framework such as data sharing, data availability, real-time handling of vast amount

of data, integration of so many components, data annotation, efficient and effective anonymisation algorithms, improving the DMT (user interface) with more functionalities, etc.

Regarding the first dimension, the MARVEL framework has enabled several mechanisms to ensure the preservation of citizens privacy. Data anonymisation of all data modalities is performed at the edge, only anonymised data is stored at the MARVEL Data Corpus, end-to-end security across all entities of the MARVEL framework, and internal device integrity of the code and data. In total, 6 safety mechanisms are exploited within MARVEL (VAD, AudioAnony, VideoAnony, EdgeSec VPN, EdgeSec TEE and deployment on edge devices for audio components, as VideoAnony is too resource-hungry to be deployed on-premise). Moreover, MARVEL fully complies with all the ethical issues related to data processing, secondary use of personal data, research activities in non-EU countries, informed consent for data processing activities, algorithmic bias, data misuse, risk of mass surveillance, anonymisation and pseudonymisation techniques, and Responsible and Trustworthy AI. To ensure ethics compliance a dedicated task T2.5 for ethics-related issues and monitoring covering all the phases of the project that is being led by the MARVEL Data Protection Manager (PN), all ethics issues are monitored and reported on multiple levels of project management and independent oversight via the MARVEL EB. One thing that is pending is to convince the external stakeholders that MARVEL does not violate the privacy of the citizens in the upcoming events with dedicated sessions and concrete examples of how this is being achieved within MARVEL.

Regarding the second dimension, MARVEL has progressed a lot since the MVP Info Day and the concerns that were raised during the event were taken into consideration. The pilots have released a significant amount of data that after being anonymised, it is stored in the MARVEL Data Corpus (0.58 PB), 31 out of the 35 components have been integrated into the 1st release of the MARVEL framework (more details in D5.4). A significant amount of data was annotated to train the AI models that were used in the 5 use cases of the first reporting period. However, it is a fact that annotation is an extremely time-consuming activity and we had to decide how much effort we would devote to this task. The DMT was significantly improved by adding more functionalities, such as (i) analytical reasoning for medium to long-term business decision-making based on queries execution over the processed AV data, (ii) data presentation and advanced visualisations that reveal hidden insights of valuable knowledge, (iii) text-annotated attention maps which will enhance video streams with textual information and indications of associated audio events, (iv) multisource, multimodal summaries, that allow users to explore and understand AV, sensor and other context-enriching data, and (v) real-time visualisations of alerts and detected events for short-term decisions and monitoring, supported by a rule-based engine (more details can be found in D4.3[17]).

The 2nd Info Day will take place in November 2022 in Malta. The feedback that we will receive will be the driving force for the final release of the MARVEL framework at M30. The results of the Info Day, meaning the feedback and the steps that the consortium undertook to address them, will be reported in D6.4 (due M36). D6.4 will also include the results from the final Info Day that will be organised before the end of the project.

[17] MARVEL D4.3: MARVEL's decision-making toolkit – initial version, 2022. https://doi.org/10.5281/zenodo.6821280

10.2.2 MARVEL Workshop organised at DataWeek 2022

Data Week 2022[18] was held both online (May 25th to June 3rd) and physically (June 8th – 9th) this year. Data Week is the spring gathering of the European Big Data and Data Driven AI research and innovation communities, where the participants share knowledge and results, discuss topics of common interest, find synergies, build new collaborations and identify new challenges and recommendations. Data Week is co-organized by the Big Data Value Association[19] and the EUHubs4Data[20] project. This year's main theme was "Towards an innovative, trusted, and fair European data economy".

MARVEL organised a workshop titled "The challenges of the extreme-scale multi-modal analytics applications"[21]. The session explored the latest EU research outcomes related to Extreme Data Analytics applications in smart cities, healthcare, robotics, and the maritime industry. Representatives from four different projects participated, namely MARVEL (represented by FORTH and ZELUS), ExaMode[22], OpenDR[23], and VesselAI[24]. The participants focused on the key challenges addressed in each sector in an effort to showcase the lessons they learned the hard way and encourage other researchers to take them into consideration. In particular, the key outcomes of the session were:

- Pay attention to ethics and privacy concerns and account for a dedicated task on this topic which is a matter of high concern for all users.
- Integration is always complex. Do not try to do everything in parallel, rather pace your integration, especially if you are dealing with E2F2C continuum.
- Reserve time for proper benchmarking and validation.
- Consider earlier versions of software solutions. They may not have the latest features in place, but they are more stable and more predictable.
- Allocate a task for data annotation since it is time-consuming and needs a lot of effort.
- Active perception using re-enforcement learning: Aggressive word shaping may be required and also consider supervised approaches.
- Interacting with clinical partners was challenging due to the Covid-19 pandemic.
- Need to be up-to-date with regards to new technologies' advancements and be able to quickly adapt/integrate them into the pipelines of the project.
- Projects may face problems with respect to the time plan that was designed during the proposal shaping. If you do not have an HPC infrastructure ready in place when the project starts then keep in mind that it will need a significant amount of time to be ready for running/training your solutions. Other options should be considered in order not to delay the development of the project.

[18] https://www.big-data-value.eu/data-week-2022/

[19] https://www.bdva.eu

[20] https://euhubs4data.eu

[21] https://www.marvel-project.eu/the-challenges-of-the-extreme-scale-multi-modal-analytics-applications/

[22] https://www.examode.eu

[23] https://opendr.eu

[24] https://vessel-ai.eu

118 people attended the session and actively participated in the discussion with questions during the last half hour of the session. As a conclusion, it was quite surprising the fact that the projects are facing similar challenges and there were also significant commonalities in the lessons learned. These lessons learned were communicated to the MARVEL consortium so that if needed the successful paths that were introduced during the event by the other projects can be followed.

11 Summary and Conclusions

This document reported an interim evaluation of the MARVEL framework from a use case point of view, including whether the components satisfy the use case technical requirements, whether end-users perceive the tools developed as useful in practice and to what extent the use cases implemented can potentially address societal challenges. The latter requirement is necessary since the use cases were motivated by societal challenges and any feedback gathered at this half-way point can be fed back into the iterative design process that the MARVEL project follows.

The framework is useful if the technical components meet the use case requirements. The MARVEL framework consists of several components that interact to deliver a dynamic E2F2C infrastructure. The feedback from the AB and the EB has helped in shaping some of the component requirements. The performance of most individual components has now been measured mostly in the lab against industrial benchmarks and in addition some components have been tested on real-world data obtained from the use cases and even under standard operation on the field. Whilst some components have achieved the intended KPIs, others require further improvement. In some cases, the combination of a limited number of components has also been tested and a number of other evaluation experiments have been planned. The end-to-end tests required for the evaluation of the framework KPIs are currently being planned over the R2 period.

Evaluating the framework from a user point of view requires the engagement of professionals who are the target users of the various use cases, i.e., transport planners/engineers and law enforcement and security personnel. For this reason, the opinion of the end-users was gathered using both formal and informal surveys for some of the use cases. The surveys included opinions on how much useful the tools developed are to fulfil their intended application. The opinions gathered will serve as feedback during the technical development of R2 period.

Understanding to what degree societal challenges are addressed is a more difficult exercise since in most cases measurements can only be carried out once a use case or a tool is widely deployed over a number of years. This requirement is impossible to achieve due to the limited timeframe of the project and therefore the evaluations can only rely on the perceived impact of the use cases. The first info day organised in Trento collected the first feedback from smart city decision-makers, which served to fine-tune the use cases prior to the R1 integration phase. Following the R1 integration phase, surveys limited to a couple use cases have questioned both personnel working in the field of interest, on efficiency and user experience, as well as the public, for their opinion on meeting societal challenges. The opinions gathered in these surveys and events will serve as feedback during the R2 period. In addition, several other events are planned for R2 such as the 2nd Info Day that will take place in November in Malta were the MARVEL framework will be presented and the goal is to collect feedback in a structured way, and more surveys with external stakeholders.

The groundwork for evaluating the business KPIs, trustworthiness KPIs and framework KPIs has been carried out and these KPIs will be evaluated at the opportune moment as the R2 period unfolds in time.

Finally, this interim evaluation report leads to the final evaluation and analysis report (D6.4) which is due at the end of the project.