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

D6.3: Demonstrator execution - final version[†]

Abstract: This deliverable provides a detailed description of the experiments' realisation following the line of D1.2 and D6.1, taking into consideration the progress of the work done from month 19 (July 2022) to month 30 (June 2023) concerning the experimental protocol alignment of Task 6.1 and the configuration of the framework and execution of real-life societal use cases of Task 6.2. The implementation of the use cases planned for the second part of the project (R2) and the variations made to the use cases implemented in the first part of the project (R1) are reported. Each use case is described in terms of framework and data flows configuration, applied multimodal and privacy-aware intelligence, execution of real social experiments, as well as demonstration of operation by reporting the experimental indicators and associated metrics for all experiments (Task 6.3).

Contractual Date of Delivery	30/06/2023
Actual Date of Delivery	21/07/2023
Deliverable Security Class	Public
Editor	*Thomas Festi (MT)*
Contributors	All *MARVEL* partners
Quality Assurance	*Nicole Bonnici (GRN)* *Adrian Muscat (GRN)* *Pawel Bratek (PSNC)*

[†] *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. *Nicole Bonnici, (GRN)*
2. *Adrian Muscat, (GRN)*
3. *Pawel Bratek, (PSNC)*

Revisions

Version	Date	By	Overview
0.4.0	21/07/2023	Thomas Festi	Addressing final comments from PC
0.3.0	10/07/2023	Thomas Festi	Close all the comments of the internal reviewers
0.2.0	30/06/2023	Thomas Festi, Nicole Bonnici, Adrian Muscat, Pawel Bratek	Include second-round partners' contributions Address comments of the internal reviewers
0.1.4	25/06/2023	Thomas Festi	First draft for internal review
0.1.3	16/06/2023	Thomas Festi	Include partners' contributions
0.0.3	01/06/2023	Thomas Festi	Rationalisation and synchronisation with D5.6
0.0.2	17/05/2023	Thomas Festi, Dragana Bajovic	Final ToC
0.0.1	10/05/2023	Thomas Festi, Dragana Bajovic, Alexandros Iosifidis	Comments on ToC
0.0.0	21/04/2023	Thomas Festi	ToC - draft version

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
AI	Artificial Intelligence
AudioAnony	GANs for audio anonymisation
AV	Audio-Visual
AVAD	Audio-Visual Anomaly Detection
AVCC	Audio-Visual Crowd Counting
AT	Audio tagging
AUC	Area Under the ROC Curve
CATFlow	Data Acquisition Framework
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
E2F2C	Edge to Fog to Cloud
EC	European Commission
EdgeSec	Security Services at the edge
ELAN	EUDICO Linguistic Annotator
FLOPS	Floating Point Operations Per Seconds
FPS	Frames Per Second
GAN	Generative Adversarial Network
GB	Gigabyte
GDPR	General Data Protection Regulation
GPU	Graphics Processing Unit
GPURegex	GPU Pattern Matching Framework
H2020	Horizon 2020 Programme
HDD	Hierarchical Data Distribution
HPC	High Performance Computing
HTTP	HyperText Transfer Protocol
HTTPS	HyperText Transfer Protocol Secure

HW	Hardware
ICT	Information and Communication Technology
IoT	Internet of Things
IP	Internet Protocol
IT	Information Technology
JSON	JavaScript Object Notation
KPI	Key Performance Indicator
M#	Month #
MAE	Mean Absolute Error
MB	Megabyte
MEMS	Micro Electro-Mechanical Systems
ML	Machine Learning
MP4	MPEG-4 Part 14 digital multimedia container format
MPEG	Moving Picture Experts Group
MQTT	Message Queuing Telemetry Transport
MVP	Minimum Viable Product
O#	Objective #
openSMILE	open-source Speech and Music Interpretation by Large-space Extraction
PC	Personal Computer
POE	Power Over Ethernet
R#	Release
R	Report
RAM	Random Access Memory
RPi	Raspberry Pi
RTSP	Real-time Streaming Protocol
sec	second
SED	Sound Event Detection
SED@Edge	Sound Event Detection at the Edge
SELD	Sound Event Localisation and Detection
SmartViz	Advanced Visualisation Toolkit
SOTA	State-of-the-Art
SSD	Solid State Drive
T#.#	Task #.#
TAD	Text Anomaly Detection
TOC	Table of Contents

UC#	Use Case
UCSD	User-Centered System Design
UI	User Interface
URL	Uniform Resource Locator
USB	Universal Serial Bus
VAD	Voice Activity Detection
VCC	Visual Crowd Counting
ViAD	Visual Anomaly Detection
VideoAnony	GANs for video anonymisation
VPN	Virtual Private Network
WAV	Waveform Audio File Format
WP#	Work Package #
Y#	Year #

Executive Summary

MARVEL aspires the convergence of a set of technologies in the areas of AI, analytics, multimodal perception, software engineering, and HPC as part of an Edge-Fog-Cloud Computing Continuum paradigm, to support data-driven real-time application workflows and decision-making in modern cities, showcasing the potential to address in an effective way societal challenges.

This deliverable is the final document of the configuration of the MARVEL framework and the execution of real-life societal experiments in smart city environments (from month 10 to month 36). In particular, it reports on the achievements of the experimental protocol alignment, the configuration of the framework, the data streams, the multimodal and privacy-aware intelligence and its execution in real-life societal experiments in each use case selected for implementation for month 30. At the same time, it discusses the changes related to the use cases realised up to month 18.

This document also provides a detailed list of indicators to be measured during the experiments in order to validate the technical, functional, and non-functional performance of the MARVEL platform. Special emphasis is placed on ensuring the alignment of the operational experiments to foster innovation in audio-visual analytics and sound recognition as well as on addressing societal and industrial requirements in the smart city domain expressed in the context of the defined use cases.

Finally, it provides a detailed report of the implementation, problems encountered, solutions developed with respect to the expected project innovations and goals, as well as an overview of the activities carried out. The specifications of the operational experiments outlined in this deliverable will be further refined and adjusted during the benchmarking and evaluation phase to enhance the platform's functionality from the perspective of the end users.

1 Introduction

1.1 Purpose and scope of this document

This deliverable presents the final implementations of WP6 – Real-life societal experiments in smart city environments. The document primarily focuses on the process of aligning the experimental protocols to ensure a seamless and appropriate implementation of the experiments based on the established protocol. Additionally, it provides updates on the progress of preparatory actions, such as developing execution time plans, defining evaluation scenarios, and selecting framework tools to be tested. It also discusses the reasons behind adapting the system modules for the trial execution.

Furthermore, the deliverable offers an updated specification outlining the iterative execution of the experiments throughout the entire duration of WP6. It acknowledges that the current version of the MARVEL framework can be refined in terms of functionality and highlights the need for further testing and validation by the final end users.

Finally, the document presents the outcomes derived from all real-life smart city experiments that have been conducted. These outputs will undergo analysis to assess the framework in terms of effectiveness, operability, usability, robustness, performance, accountability, transparency, and privacy awareness.

1.2 Contribution to WP6 and project objectives

This document is the main output of Task 6.1 – Experimental protocol alignment and Task 6.2 – Configuration of the framework and execution of real-life societal experiments. In addition, it provides an overview of the ongoing work under Task 6.3 – Evaluation and Impact analysis.

The deliverable is directly related to the achievement of MARVEL Objective 4: *"Realise societal opportunities in a smart city environment by validating tools and techniques in real-world settings"*. As modern cities face numerous societal challenges, it is critical for technologies like Big Data, IoT and Edge/Fog/Cloud computing to offer solutions that increase citizens' wellness and well-being. However, it is critical also to encapsulate the complexity of a city and support accurate, cross-scale, and in-time predictions across different application scenarios.

This involves:

- ensuring smooth and adequate execution of the experiments according to the experimental protocol (T6.1 and T6.2);
- demonstrating how the MARVEL framework can quickly and effectively aggregate, process and visualise extremely-large-scale audio-visual data at the edge, fog, and cloud (T6.1 and T6.2);
- fostering innovation in audio-visual analytics and sound recognition and addressing societal and industrial requirements in the smart city domain (T6.1, T6.2, and T6.3);
- providing structured feedback, from both the data providers and the technology owners to the development process (T6.3);
- assessing project impact to drive actions for the framework's long-term sustainability (T6.3 and related to T7.5).

The work performed under the WP6 – Real-life societal experiments in smart cities environment contributes to achieving the following project-related KPIs:

- KPI-O4-E1-1: More than 10 trial cases to showcase framework's capabilities;

- KPI-O4-E2-1: Identify at least 20 dependent and independent verification and validation variables for the system;
- KPI-O4-E3-1: Execute the trial cases in at least two real life smart city environments.

Finally, the activities carried out contribute to MS7 – MARVEL integrated version (2nd release).

1.3 Relation to other WPs and deliverables

This deliverable synthesises the final results of all WP6 activities, and as such, there is a close interrelation between this deliverable and all the tasks in the current WP. Furthermore, there is also a strong dependency between this deliverable and many other WPs and as all scientific and technical developments of the project directly influence the successful implementation of the use cases:

- WP1: Setting the scene: Project set up:
  - T1.3: Experimental protocol - real life societal trial cases in smart cities environments and D1.2 – MARVEL's experimental protocol.
- WP2: MARVEL multimodal data Corpus-as-a-Service for smart cities:
  - T2.1: Collection and analysis of MARVEL experimental distributed data assets and D2.1 – Collection and analysis of experimental data;
  - T2.2: Data management and distribution, D2.2 – Management and distribution Toolkit – initial version and D2.4 – Management and distribution Toolkit – final version.
- WP3: AI-based distributed algorithms for multimodal perception and situational awareness:
  - T3.2: MARVEL's personalised federated learning realisation for extreme-scale analytics and D3.4 – MARVEL's federated learning realisation;
  - T3.3: Multimodal audio-visual intelligence, D3.1 – Multimodal and privacy-aware audio-visual intelligence – initial version and D3.5 – Multimodal and privacy-aware audio-visual intelligence – final version;
  - T3.4: Adaptive E2F2C distribution and optimisation of AI tasks, T3.5: Edge-optimal ML/DL deployment for multimodal processing, D3.2 – Efficient deployment of AI-optimised ML/DL models – initial version and D3.6 - Efficient deployment of AI-optimised ML/DL models – final version.
- WP4: MARVEL E2F2C distributed ubiquitous computing framework:
  - T4.1: Optimised audio capturing through MEMS devices and T4.2: openSMILE platform for audio-visual analysis and voice anonymisation, D4.1 – Optimal audio-visual capturing, analysis and voice anonymisation – initial version and D4.4 – Optimal audio-visual capturing, analysis and voice anonymisation.
- WP5: Infrastructure Management and Integration, mainly regarding:
  - T5.1: HPC infrastructure, D5.3 – HPC infrastructure and resource management for audio-visual data analytics – initial version and D5.8 - HPC infrastructure and resource management for audio-visual data analytics – final version;
  - T5.3: Continuous integration towards MARVEL's framework realisation, D5.1 – MARVEL Minimum Viable Product, D5.4 – MARVEL Integrated framework – initial version and D5.6 – MARVEL Integrated framework – final version.

1.4 Structure of the document

The structure of this document is as follows:

- Section 2 – The experimental protocol definition is optimised, considering the outcome of the MVP deployment at M12 and the first MARVEL prototype released at M18, in terms of architecture, component interaction; any new insight derived from the data collections; hardware procurement and deployment.
- Section 3 (GRN1: Safer Roads), Section 4 (GRN2: Road User Behaviour), Section 8 (MT2: Detecting Criminal and Anti-Social Behaviours), Section 10 (MT4: Analysis of a Specific Area), and Section 12 (UNS2: Localising Audio Events in Crowds) – The framework configuration and data streams; the multimodal and privacy-aware intelligence applied; the demonstration of the implementation realised and work carried out are presented.
- Section 5 (GRN3: Traffic Conditions and Anomalous Events), Section 6 (GRN4: Junction Traffic Trajectory Collection), Section 7 (MT1: Monitoring of Crowded Areas), Section 9 (MT3: Monitoring of Parking Places), and Section 11 (UNS1: Drone Experiment) – The report of the updates compared to the first project phase (M18) is introduced.
- Section 13 – The summary and conclusions of the document are included.

2 Experimental protocol alignment

Deliverable D1.2[1], released at M8, provided a first definition of the experimental protocol and the benchmarking strategies for the MARVEL project, considering the framework as a whole as well as each individual component. The experimental protocol has been then revised in D6.1[2], particularly in terms of component KPIs, execution plan and use case goals, for the five use cases involved in the first MARVEL prototype released at M18. This revision was necessary to account for practical implementation and deployment issues as well as for the evolution of the technological components.

The same process was applied for the second release of the MARVEL project (M30), with a particular focus on the five new use cases and considering the experience in the realisation of the R1. More details about the alignment procedure and the scope of task T6.1 are available in D1.2 and D6.1.

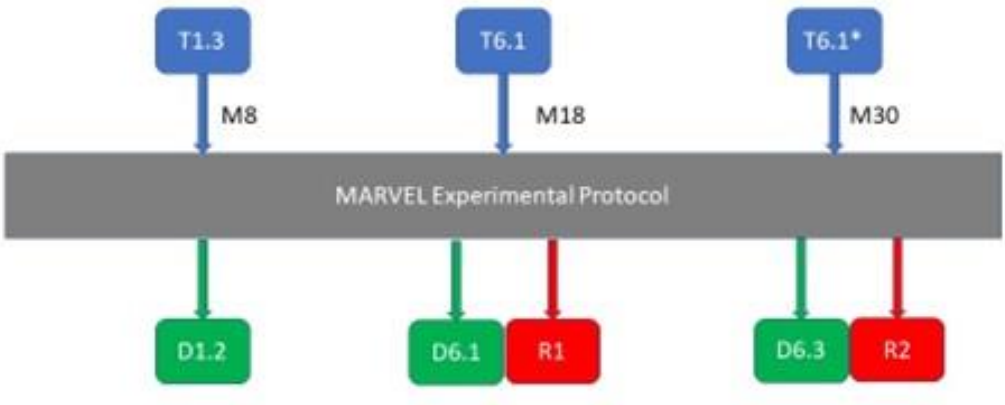

Figure 1. Evolution of the MARVEL experimental protocol

2.1 Scope of Task 6.1

T6.1 was originally planned to end at M18. However, it was considered necessary this task to be extended until M30 to better align the experimental protocol taking into consideration the five new use cases deployed in the R2. Details about the T6.1 scope and goals are available in D6.1. As a result of this second round of the experimental protocol alignment:

- new components were introduced with related KPIs;
- asset KPIs were slightly revised;
- one new use case was defined (UNS2) with related KPIs and requirements;
- functional and non-functional KPIs of GRN1 and GRN2 were revised.

Note that the one of the objectives of T6.1 which involves the final selection of framework tools to be tested, is comprehensively documented in D5.6[3]. This deliverable pertains to the final architecture of the MARVEL prototype for each use case, outlining the specific tools chosen and their integration within the overall framework.

[1] "D1.2 - MARVEL's experimental protocol," Project MARVEL, 2021. Confidential.

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

[3] "D5.6 - MARVEL Integrated framework – final version," Project MARVEL. To be released.

2.2 Update and Revision of the execution plan

This section revises and updates the execution time plans for the three pilots with respect to what was planned in D1.2 and revised in D6.1.

2.2.1 Greenroads time plan

The time plan for GRN use cases, which was previously outlined in D6.1, has been updated and is presented in Table 1.

Table 1: GRN time plan

Phase	Activity	Period	Status
Baseline	Definition of the use cases	M6	Completed
	First data release	M8	Completed
Innovation	HW procurement	M8-M12	Completed
	HW computing devices installation	M10-M22	Completed
	HW sensing devices installation	M10-M22	Completed
	First release of the MARVEL's infrastructure	M12	Completed
	Minimum Viable Product	M12	Completed
	MARVEL prototype 1st version	M18	Completed
Experimentation	Internal Evaluation of the R1	M22	Completed
	Revision of the data flow	M18-M24	Completed
	Execute field trials	M20-M30	Ongoing
Consolidation	Release the multimodal audio-visual Data Corpus	M22-M30	Ongoing
	Fix issues	M22-M30	Ongoing
	Develop a business plan	M22-M30	Ongoing

No deviations are reported with respect to the time plan reported in D6.1. The table was updated with respect to the current status of the tasks.

2.2.2 Municipality of Trento time plan

The time plan for MT use cases, which was previously outlined in D6.1, has been updated and is presented in Table 2.

Table 2: MT time plan

Phase	Activity	Period	Status
Baseline	Definition of the use cases	M6	Completed
	First data release	M8-M12	Completed
Innovation	HW purchasing and installation	M8-M26	Completed
	First release of the MARVEL's infrastructure	M18	Completed
	MARVEL prototype 1st version	M18	Completed

Experimentation	Internal Evaluation of the R1	M22	Completed
	Revision of the data flow	M18-M24	Completed
	Execute field trials	M20-M30	Ongoing
Consolidation	Release the multimodal audio-visual Data Corpus	M22-M30	Ongoing
	Fix issues	M22-M30	Ongoing
	Develop a business plan	M22-M30	Ongoing

No deviations are reported with respect to the time plan reported in D6.1. The table was updated with respect to the current status of the tasks.

2.2.3 Novi Sad time plan

The time plan for UNS use cases, which was previously outlined in D6.1, has been updated and is presented in Table 3.

Table 3: UNS time plan

Phase	Activity	Period	Status
Baseline	Definition of the use cases	M8	Completed
	Redefinition of the UNS2 use case	M26	Completed
	HW testing	M7-M27	Completed
Innovation	Recording app development	M10	Completed
	Databases recording	M12-M28	Completed
	First data release	M14	Completed
Experimentation	Algorithm development	M20-M30	Ongoing
	MARVEL prototype 1st version	M18	Completed
	Execute field trials	M18-M30	Ongoing
Consolidation	Release of recorded databases	M22-M29	Completed
	Fix issues	M22-M30	Ongoing
Use case evaluation	Future exploitation plans	M30-M36	Ongoing

UNS had to revise the time plan reported in D6.1 as the Audio-visual emotion recognition use case (UNS2 use case) was redefined. The previous UNS2 use case envisioned an application where audio-visual data is gathered by a close-by camera and microphone (e.g., a mobile phone) and with the assumption that the identity information of the recorded individual was available (e.g., a service at a bank counter). This use case had then been identified as challenging from an ethical point of view. However, UNS as a data provider had already made significant progress regarding the abandoned use case including an initial small-scale database recording through a staged-recording process, which was reported in Table 13. In terms of technical work, UNS had also developed the privacy-preserving audio-visual emotion recognition component, including an accompanying federated learning component to address the cases where the data

is held by different parties, which has been reported in D3.5[4]. However, to address the ethical concerns which arose a need to develop a novel component (which was not addressed in R1) and an accompanying use case was identified. The nature of the UNS2 use case was fully revised and it was redefined to focus on SELD. Detailed time plan of the (new) UNS2 use case can be found in Table 12. For details on privacy preservation of the UNS2 use case, please check Section 1.1.1.12.2.1.3.

An analysis of ethical challenges of emotion recognition applications is provided in Appendix.

UNS has allocated at the end of the use case to conduct a comprehensive evaluation and explore future exploitation opportunities. Within this phase, we will perform evaluation of the use cases by measuring the use case defined KPIs and also explore possibilities for collaboration with the city of Novi Sad for the future smart city Novi Sad pilot.

2.3 R1 use cases

Five use cases were implemented in the first release of the MARVEL prototype at M18: GRN3, GRN4, MT1, MT3, and UNS1.

Based on the evaluation results, practical aspects, infrastructure, and new evolutions of the technological components, these five use cases have been refined with respect to what was reported in D6.1. The following sections describe these changes.

2.3.1 GRN3: Traffic Conditions and Anomalous Events

The use case is useful to monitor traffic conditions and detect 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 interpret the data and take necessary actions.

The experimental protocol for GRN3 was not changed since reporting in D6.1. The main changes concern updates to the infrastructure which were intended to improve the performance of the use case. Some updates to the user interface were also requested to improve the user experience.

2.3.2 GRN4: Junction Traffic Trajectory Collection

Junction Traffic Trajectory collection 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. 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.

[4] "D3.5 - Multimodal and privacy-aware audio-visual intelligence – final version," Project MARVEL, 2023. https://doi.org/10.5281/zenodo.8147164

The GRN4 experimental protocol is the same as the one described in D6.1. The main changes were infrastructure updates designed to improve the performance of the use case. Some user interface updates were also requested to improve the user experience.

2.3.3 MT1: Monitoring of Crowded Areas

Monitoring of Crowded Areas aims to identify relevant areas with significant crowd presence for various reasons, such as unusual movements or suspicious behaviours. One such area is Piazza Fiera, a square that hosts the annual "Christmas Markets" in Trento. These markets attract thousands of visitors. The increased crowd density can lead to an upsurge in thefts and assaults, as well as the potential need for medical assistance for individuals feeling unwell or fainting. Another location requiring close monitoring is Piazza Duomo, a square that hosts a weekly market in the city centre. Similar to the Christmas Markets, this area is prone to crowding.

The experimental protocol for MT1 has remained unchanged since its reporting in D6.1. However, notable updates were made to the user interface to enhance the overall user experience. These changes were specifically requested to optimize user interactions and ensure a more seamless and enjoyable experience throughout the experiment.

2.3.4 MT3: Monitoring of Parking Places

Monitoring of Parking Place use case concerns audio-visual monitoring of a parking place, including detection of cars out of the parking slots, car damages, car 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 system will analyse the audio-visual data to detect potential issues that may refer to the scenarios described above. The aim of the use case is to detect anomalies, the timeline distribution of the vehicles in the parking, the total number of vehicles, the clustering of vehicles and/or events, as well as the information on detections observed.

The experimental protocol for MT3 was not changed since reporting in D6.1. The main changes concern updates to the infrastructure, which was intended to improve the performance of the use case. Some updates to the user interface were also requested to improve the user experience.

2.3.5 UNS1: Drone Experiment

Drone Experiment performs the monitoring and surveillance of large public events and the behaviour of the crowds through the utilisation of drones. There were no changes regarding UNS1 use case experimental protocol. Several refinements have been made to the user interface:

- the way crowd counting maps are handled from the technical view point;
- download option of a snapshot of the system state, logs, etc. as evidence of what has occurred in the system, which can be shown to the police or local authorities was introduced.

UNS also reported that for continuous use for a year, the RPI device malfunctioned, but it was changed with a new RPI v4 device. Finally, infrastructure was updated including EdgeSec TEE.

2.4 R2 use cases

This section describes the five new use cases deployed in R2, focusing in particular on the user stories and evaluation scenarios. These have been initially introduced in D1.2 but are formally

defined here, given the experience of R1 and the technology and infrastructure developments achieved during the project.

2.4.1 GRN1: Safer Roads

This use case addresses the need to increase safety on urban roads for vulnerable road users, with the aim of encouraging the uptake of active travel modes in Malta. More specifically, this use case targets cycling and walking. Malta has witnessed a significant effort, from both the authorities and the bicycle commuting lobby, in encouraging cycling and walking, mainly through infrastructural changes. The use case takes this effort further and aims at detecting cyclists, including e-bikes and pedestrians, exiting a junction and alert car and motorised-vehicle drivers of their presence via variable message signs with the hope that car drivers take greater care and concentrate more in such circumstances.

In addition, detecting cyclists is a particularly interesting task in low visibility conditions because it is both more dangerous for these entities and more challenging from a technology point of view.

This use case should contribute towards an increase in the perceived safety on the roads and will therefore encourage commuters to consider cycling as an alternative mode of transportation. To determine the impact of this use case, surveys to gauge citizens' perceptions of safety with this device will be conducted.

Execution time plan

The execution time plan up to the M36 and evaluation scenarios are reported in Table 4 for GRN1.

Table 4: GRN1 execution time plan

Phase	Activity	Period	Status
Baseline	Definition of the use cases	M6	Completed
	First data release	M8	Completed
Innovation	HW procurement	M8-M12	Completed
	HW computing devices installation Equipment – released on M26: • *Apollo Dev Kit delivered to AU for ease of development;* • *Apollo dev Kit Connected to network for GRN;* • *LED sign and Arduino controller connected and tested by GRN;* • *Camera Connected to Jetson via RTSP;* • *Anonymisation on Fog connected.*	M10 – M26	Completed
	Availability of Training Data to fine-tune models – released on M26: • *GRN provided trained model for YOLO-SED component;* • *GRN provided labelled data for the SED dataset;* • *GRN is providing a labelled multimodal dataset for YOLO-SED.* User interface – released on M26: • *Implemented in SmartViz.*	M26	Ongoing

	Testing of use case – released on M26: • *Edge-to-Edge testing is in progress, components tested individually.* Deployment of use case – released on M27: • *This is a collective effort between all the partners involved in the use cases.*		
Experimentation	Internal Evaluation of the R2	M27	Not Started
	Execute field trials	M24-M27	Ongoing

Final Evaluation Scenarios

Two evaluation scenarios have been defined for this use case related to two use case KPIs as reported in Table 5.

Table 5: Evaluation scenario and relevant KPIs for the GRN1

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	Improvement in detection rate due to multimodal detection	**GRN-KPI1: Detect Vulnerable Road Users (VRUs) at any time of day, including low-light conditions**
Observing the time taken to detect an anomaly through measuring the system's latency	The aim is to detect anomalies 2 seconds after the start of the event	**GRN-KPI2: low latency (time between detecting the cyclist and informing the road users)**

Modifications to the E/F/C infrastructure

The GRN Infrastructure provided for the R2 integration of GRN1 consists of one IP camera in Mgarr (a rural town in the northwest region).

GRN has deployed two Jetson Apollo Dev Kits which can connect directly to the Mgarr Camera. One of the GRN Devices was deployed in Malta for testing whilst another was deployed in AU for development purposes. The GRN Fog layer (provided by FORTH) was also used for this use case, mainly to anonymise the stream before it is displayed on the user interface.

Datasets

GRN provided support and resources in multiple ways to the development of the YOLO-SED component. In more detail:

- GRN provided a dataset for the development of SED throughout the project with recent snippets chosen to balance the dataset for rare classes such as bus, motorcycle and bicycle;
- for the YOLO-based part component, GRN provided the pre-trained model weights from CATFlow;
- GRN produced a combined multimodal dataset to test the YOLO-SED component.

2.4.2 GRN2: Road User Behaviour

This use case addresses the need to monitor the behaviour of road users at a junction. This use case demonstrates tools that are useful in law enforcement and education campaigns targeting responsible driving, cycling, and other uses of the roads. Malta has experienced fast changes in the transport landscape, to which human response often lags behind technical progress. Educational campaigns are one way to fill in the gap and have been shown to be effective in the past. This use case involves the classification of actions into a spectrum of examples demonstrating good to bad behaviour. This use case will not be implementing the latter campaigns or policies; however, it could be tried in different places and its output could be observed. Surveys will be used to find how this tool will be able to help local authorities.

Examples of actions include the way pedestrians cross over the intended crossings, whether cyclists dismount at pedestrian crossings, and whether car drivers stop in the delineated zone at junctions. The system will be able to count the number of times bad behaviour is detected before and after the execution of education campaigns or policy changes.

The data is then collected by the UI, following which a local authority can compare the data before and after an educational campaign or access the impact of a recent installation of a traffic calming measure. The impact of this use case will be measured through interviews with authorities or other third parties to determine if this tool will help them evaluate their campaigns and policies.

Execution time plan

The execution time plan, up to the M36, and evaluation scenarios for GRN2 are reported in Table 6. These plans were carefully outlined before the commencement of the integration process and were diligently adhered to throughout the course of the project.

Table 6: GRN2 execution time plan

Phase	Activity	Period	Status
Baseline	Definition of the use cases	M6	Completed
	First data release	M8	Completed
Innovation	HW procurement	M8-M12	Completed
	HW computing devices installation Equipment – released on M26: • *GRN Edge 1 improved internet connection;* • *GRN Edge 2 deployment;* • *3 GRN IP cameras maintained for this Use Case;* • *GRN Fog 2 deployed.*	M10 – M26	Completed
	Availability of Training Data to fine-tune models – released in M26: • *GRN provided training dataset for the TAD model;* • *GRN provided a small scale dataset for the SED Horn detections testing.* User interface – released on M26: • *Implemented in SmartViz.* Testing of use case – released on M26: • *Edge-to-Edge testing is in progress, components tested individually.* Deployment of use case – released on M27: • *This is a collective effort between all the partners involved in the use cases.*	M26	Ongoing
	Internal Evaluation of the R2	M27	Not Started

Experimentation	Execute field trials	M24-M27	Ongoing

Final Evaluation Scenarios

The evaluation scenario defined for this use case is reported in Table 7.

Table 7: Evaluation scenarios and relevant KPIs for the GRN2

Evaluation Scenario	Target	Relevant KPI
The evaluation scenario includes testing the various AI models to confirm the detection of at least four improper behaviours	Four improper behaviours detected	**GRN-KPI3: Automatically detect and label or quantify actions that determine driver behaviour**

Modifications to the E/F/C infrastructure

The actions to finalise the infrastructure included:

- the connection of GRN Edge 1 to a robust internet connection;
- the deployment of GRN Edge 2 to anonymise the video at a closer location to the data acquisition source;
- upgrading GRN Fog 2 to increase the computational power available;
- the use of three IP cameras.

Datasets

GRN provided a training dataset for the TAD model and a small-scale dataset for the SED Horn detections testing.

2.4.3 MT2: Detecting Criminal and Anti-Social Behaviours

The objective is to monitor specific areas for the purpose of identifying criminal or anti-social activities. The MARVEL framework was implemented to detect potentially dangerous situations. These situations include gatherings, robberies, aggressions, and drug trafficking. The system analyses the visual and audio data streams of the designated location and promptly alerts the local police operational centre, enabling them to dispatch a team to the scene. Additionally, the streams are saved on the local police server for future investigations.

The real-time analysis covers a daily timeframe. In the latter scenario, the data will be retained for seven days as per the privacy regulations outlined in the GDPR 2016/679. After this period, the data will be deleted unless the police receive specific requests for timely investigations.

To summarise, the system aims to monitor and identify bothersome gangs, instances of aggression or robbery, gang fights, and drug dealing by analysing audio and video data in real-time or from recordings. The information is used to alert the local police and stored for a limited period, adhering to privacy regulations.

Execution time plan

The execution time plan, up to the M36, and evaluation scenarios for MT2 are reported in Table 8. These plans were carefully outlined before the commencement of the integration process and were diligently adhered to throughout the course of the project.

Table 8: MT2 execution time plan

Phase	Activity	Period	Status
Baseline	Definition of the use cases	M6	Completed
	First data release	M8-M12	Completed
Innovation	HW procurement	M8-M24	Completed
	HW computing devices installation	M15–M26	Completed
	HW sensing devices installation: • *Final installation of microphones and RPIs in all the use cases*	M15–M26	Completed
	First release of MARVEL's infrastructure • *Due to the GDPR constraints, MT decided at M16 that edge tier will not be part of the Kubernetes cluster*	M18	Completed
	Availability of Training Data to fine-tune models – released from M26 to M29: • *MT provided annotated dataset for the training of AAC, AVAD GPURegex and SED components.* User interface – released on M28: • *Implemented in SmartViz.* Testing of use case – released in M28: • *End-to-End testing is in progress, components tested individually.*	M12	Completed
Experimentation	Internal Evaluation of the R2	M27	Not Started
	Execute field trials	M24-M30	Ongoing

Final Evaluation Scenarios

The evaluation scenario defined for this use case is reported in Table 9.

Table 9: Evaluation scenarios and relevant KPIs for the MT2

Evaluation Scenario	Target	Relevant KPI
The evaluation scenario involves observing the reduction in reaction time and taking appropriate intervention measures from the time of detection of the above events	The aim is to react to the detected anomalies in 5 minutes from the start of the anomalous event	**MT-KPI3: Reaction time in case of issues**
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 improves at least 50% the detection of anomalous events	**MT-KPI4: Increase the detection of targeted events**

Modifications to the E/F/C infrastructure

The MT2 infrastructure consists of:

- two IP cameras in Piazza Santa Maria Maggiore;
- two Microphones, IFAG-MEMS, in Piazza Santa Maria Maggiore;
- two Raspberry Pi 4B – for audio elaboration – at Piazza Santa Maria Maggiore.

The upload function ensures secure transmission via VPN access between MT and FBK. This process allows raw audio-video data to be seamlessly sent to the data lake at FBK, ensuring the privacy and integrity of the information being transferred.

FBK provides the Fog tier for the MT use cases. To adhere to the constraints outlined in the agreement, granting FBK access to the raw data from MT's sensors, and to meet the demands of the MARVdash Kubernetes cluster, FBK deploys two workstations, each equipped with powerful GPU processing capabilities.

Datasets

MT provided a training dataset for the AAC, GPURegex, SED, AT, and AVAD components.

2.4.4 MT4: Analysis of a Specific Area

The Municipality of Trento aims to enhance its decision-making process by monitoring key locations within the city. To achieve this, the MARVEL framework helps in counting people, cars, buses, taxis, and bikes, as well as calculating their trajectories and identifying noteworthy events during specific timeframes or throughout the day to facilitate effective decision-making.

One identified area of interest is the vicinity of the Trento train station, encompassing the road and a portion of Piazza Dante, spanning from the traffic lights on Via Dogana to the traffic lights on Via Pozzo. This scenario is likely to be integrated into the larger project called the "Smart City Control Room," which is being launched in the Municipality of Trento. This initiative aims to gather the necessary data for formulating and monitoring sustainable mobility plans and energy transition actions in the urban area.

The audio-video analysis will be performed in real-time and on recorded data stored on the servers of the Local Police.

The data collected in this use case will allow to:

- create a searchable database that will provide insights into the "detection of habits" within a specific area of the town, enabling long-term decision-making support for public authorities. The MT managers will utilise the MARVEL framework, and officers will collaborate with policymakers to analyse the system's results;
- improve the efficiency regarding urban planning, specifically in terms of traffic management and city planning. The MT managers will leverage the MARVEL framework, and officers will collaborate with policymakers to analyse the outcomes. This enhanced efficiency will contribute to more effective traffic management and urban planning processes.

Execution time plan.

The execution time plan, up to the M36, and evaluation scenarios for MT2 are reported in Table 10. These plans were carefully outlined before the commencement of the integration process and were diligently adhered to throughout the course of the project.

Table 10: MT4 execution time plan

Phase	Activity	Period	Status
Baseline	Definition of the use cases	M6	Completed

	First data release	M8-M12	Completed
Innovation	HW procurement	M8-M24	Completed
	HW computing devices installation	M15–M26	Completed
	HW sensing devices installation • *Final installation of microphones and RPIs in all the use cases.*	M15–M26	Completed
	First release of MARVEL's infrastructure • *Due to the GDPR constraints, MT decided at M16 that edge tier will not be part of the Kubernetes cluster.*	M18	Completed
	Availability of Training Data to fine-tune models – released from M26 to M29: • *MT provided annotated dataset for the training of AVAD, CATFlow and SED components.* User interface – released on M28: • *Implemented in SmartViz.* Testing of use case – released on M28: • *Edge-to-Edge testing is in progress, components tested individually.*	M12	Completed
Experimentation	Internal Evaluation of the R2	M27	Not Started
	Execute field trials	M24-M30	Ongoing

Final Evaluation Scenarios

The evaluation scenario defined for this use case is reported in Table 11.

Table 11: Evaluation scenarios and relevant KPIs for the MT4

Evaluation Scenario	Target	Relevant KPI
The evaluation scenario involves observing the timeline distribution of the vehicles, the total number of vehicles, the statistics of vehicles and/or events, the trajectories by traffic entities, the total number of persons and other security-related events so the Municipality can increase the management of traffic and security near the train station	The aim is to recognise 50% of mobility patterns and recurrent dangerous events detected	**MT-KPI7: collection of trajectories and snippets of anomalous events related to mobility as well as other security-related events**
The evaluation scenario includes testing the searchable database that will provide insights enabling long-term decision-making support for public authorities	The goal is the reduction of travel time through the city centre and the decrease in urban and mobility planning time	**MT-KPI8: Increased efficiency in the urban planning**

Modifications to the E/F/C infrastructure

The MT4 infrastructure consists of:

- four IP cameras in the vicinity of the Trento train station (one in Piazza Dante, two in Via Dogana and one in Via Pozzo);
- two Microphones IFAG-MEMS in the vicinity of the Trento train station (one in Piazza Dante and one in Piazza Dante);
- two Raspberry Pi 4B – for audio elaboration – in the vicinity of the Trento train station (one in Piazza Dante and one in Piazza Dante).

The upload function is a secure transmission over VPN access between MT and FBK in which raw audio-video will be sent to the data lake in FBK.

FBK provides the Fog tier for the MT use cases. 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.

Datasets

MT provided a training dataset for the AVAD and SED components.

2.4.5 UNS2: Localising Audio Events in Crowds

Ensuring safety in rapidly growing urban city areas and public events is a challenging task, which requires a quick response in the case of anomalous events. Commonly, static cameras are being used for monitoring. In the UNS1, the application of drone cameras and MEMS microphones for monitoring was analysed. However, video monitoring could not help much in the cases of low visibility (for example during the night) or insufficient number of cameras which is likely to happen when cameras are fixed in position. For that reason, in this use case, we analyse the potential of applying microphone array boards for monitoring public events. Microphone arrays can be used for detecting target sound events and finding the direction of the sound propagation, which could help to localise anomalous events in a crowd. Such a system could quickly detect accidents or other kinds of anomalous events.

Execution time plan

The execution time plan and evaluation scenarios for UNS2 are reported in Table 12. Note that, as mentioned above, this use case was introduced at the end of the second year of the project as a replacement of the original UNS2 use case (see D1.2). Therefore the execution time plan starts from M26.

Table 12: UNS2 execution time plan

Phase	Activity	Period	Status
Baseline	Redefinition of the use case	M26	Completed
	IFAG AudioHub Nano 8ch microphone board testing	M27	Completed
	Indoor data recording	M27	Completed
Innovation	Designing hardware setup for outdoor data recording	M27	Completed
	Creating a set of predefined audio events for outdoor dataset recording	M26 – M28	Completed
	Database recording – outdoor staged recording	M28	Completed
	Data release for SELD component training and testing	M28	Completed
	User interface – first draft	M29	Completed
	User interface – release	M30	Ongoing

Experimentation	Testing of the use case	M29-M30	Ongoing
	Deployment of the use case	M29-M30	Ongoing
	Internal Evaluation of the R2	M30	Not Started
	Execute field trials	M28-M30	Ongoing
Use case evaluation	Future exploitation plans	M30-M36	Ongoing

Before redefining the UNS2 use case in M26, as it is described in Section 12 of this document, there were several actions taken within the execution of the previously defined UNS2 Audio-visual emotion recognition use case. Executed actions are summarised in Table 13.

Table 13: Completed items from the execution time plan of abandoned UNS2: Audio-visual Emotion Recognition

Phase	Activity	Period	Status
Baseline	Definition of the use case	M8	Completed
Innovation	Android app development for dataset recording	M11-M12	Completed
	Scenario definition and start of staged recordings	M13	Completed
	Database recording: 4 persons recorded audio-visual data according to the defined scenario	M18	Completed

Final Evaluation Scenarios

The evaluation scenario defined for this use case is reported in Table 14.

Table 14: Evaluation scenarios and relevant KPIs for the UNS2

Evaluation Scenario	Target	Relevant KPI
Localisation of target events and alerting event organisers about them	10% relative improvement on metrics in DCASE setup (dataset, cross-validation setup)	**UNS-KPI5: Accuracy Location-dependent error rate and F1-score**
Detection of target events and alerting event organisers about them	10% relative improvement on metrics in DCASE setup (dataset, cross-validation setup)	**UNS-KPI6: Accuracy Classification-dependent localisation error and recall**
Monitoring (e.g., by security crew) of streaming data	Decrease in time comparing to the time needed to human operator to localise and detect target event from audio	**UNS-KPI7: Reaction time**

Modifications to the E/F/C infrastructure

The infrastructure of the UNS2 use case covers all three layers: Edge, Fog, and Cloud. IFAG 8 channel AudioHub Nano microphone boards are used at the edge for data recording and streaming data to the UNS Edge 3, which is a laptop. The SELD component is deployed at the same device where inference is performed. The UNS Fog server is used for deploying AVRegistry, DatAna Fog, and StreamHandler, whereas the Cloud server is used for visualisation and SELD training.

Datasets

UNS has prepared an 8-channel audio dataset, which consists of selected target events from the FSD50K database. The following target classes are selected: gunshot, shatter, and boom. These target events are extracted and combined with samples from the chatter class in order to simulate such target events as anomalous events in crowds. Further details on this dataset and annotations can be found in Section 1.1.1.12.2.1.1 and Section 1.1.1.12.2.1.2 of this document.

2.5 Development of MARVEL framework for selected use cases for M30

2.5.1 Revision of MARVEL conceptual architecture

The revision of MARVEL conceptual architecture due to the applied revisions of the use cases and the components is reported in D5.6.

2.5.2 Components

The new components with related KPIs are reported in this section. All other components are available in D5.6.

2.5.2.1 YOLOSED

YOLOSED is a new component that detects vulnerable users on the street in order to notify incoming drivers to take special care. This component utilises a YOLO object detector to detect bicycles and pedestrians and utilises SED to enhance the prediction accuracy for bicycles, which are notoriously hard to discern from motorcycles. The component contributes to the project KPI KPI-O2-E3-3.

2.5.2.2 RBAD

The Rule-Based Anomaly Detection (RBAD) is a new lightweight, rule-based anomaly detector. It is an add-on to the CATFLOW component and utilises mappings, rules and public information, like bus schedules, to detect user-defined anomalies. It contributes to KPI-O2-E3-3.

2.5.2.3 VAD+AudioAnony

The component has been heavily modified to handle the 8-channel input from the new IFAG MEMS microphone array. Nevertheless, the functionalities remain unaltered and a revision is not required.

2.5.3 MARVEL framework

The modifications applied to the use cases and to the components do not have implications on the evaluation of the MARVEL framework.

2.6 KPI revision

In the process towards the development and deployment of the R2, some KPIs had to be revised in order to account for the new/redefined operational conditions, for example when new components or new use cases were introduced. This section reports the revised KPIs for both assets and use cases. Each subsection refers to one or more KPI tables reported in either D1.2 or D6.1, integrating them with the new KPI definitions. For each KPI modification, a brief explanation is provided. Note that Section 2.6.4 reports also the KPIs of the two new components, RBAD and YOLOSED.

2.6.1 Sensing and Perception

There was no need to review the KPIs of the components in this field after the reporting in D1.2 and D6.1.

2.6.2 Security, Privacy and Data Protection

The KPIs for the AudioAnony component have been revised, as reported in Table 15.

Table 15: Revised asset specific KPIs for the "Security, privacy, and data protection" subsystem

Asset	KPI	Metric	Baseline SOTA	Datasets / Benchmarks	Expected result	Relevant project KPIs
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% WER improvement over baseline. 20% SED improvement over baseline	

2.6.2.1 AudioAnony

One of the AudioAnony KPIs needed to be refined as the metric originally considered was found to be not suitable. The previous definition "20% PESQ improvement over baseline" is replaced by "20% WER improvement over baseline".

The motivation is that PESQ measures the similarity with respect to a target speech signal, which, in this case is the input speech segment. Since the goal of AudioAnony is to modify the speaker signature, a good anonymisation tool will always lead to a smaller PESQ. Using WER instead, one can measure the amount of distortions, introduced eventually by the anonymisation component, which would compromise the speech recognition performance of a state-of-the-art solution.

2.6.3 Data Management and Distribution

There was no need to review the KPIs of the components in this field after the reporting in D1.2 and D6.1.

2.6.4 Audio Visual and Multimodal AI

The KPIs for the AVAD, ViAD, AVCC, and VCC components have been revised, as reported in Table 16. In addition, the table reports the KPIs for the two new components: RBAD and YOLO-SED.

Table 16: Asset specific KPIs for the "Audio visual and multimodal AI" subsystem – addendum with respect to D1.2

Asset	KPI	Metric	Baseline SOTA	Datasets / Benchmarks	Expected result	Relevant project KPIs
AVAD	Accuracy	Area Under	Multi-Modal	MARVEL-Malta Audio Visual	80% on MAVAD	KPI-O2-E2-1

		ROC Curve	Anomaly Detection by Using Audio and Visual Cues	Anomaly Dataset (MAVAD)		KPI-O2-E3-1 KPI-O2-E3-2 KPI-O2-E3-3 iKPI-3-3
	Speed	FLOPS			30% speedup while retaining 90% accuracy	
AVCC	Accuracy	MAE MSE	AudioCSRNet	DISCO	15.00 MAE	KPI-O2-E3-1 KPI-O2-E3-2 KPI-O2-E3-3
	Speed	FLOPS			30% speedup while retaining 90% accuracy	
ViAD	Accuracy	Area Under ROC Curve	Anomaly ARnet	UCSD pedestrian dataset	90% AUC	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	
VCC	Accuracy	MAE	SASNET	DISCO	11.00 MAE on average for sub-regions of the image	KPI-O2-E3-1 KPI-O2-E3-2 KPI-O2-E3-3
	Speed	FLOPS			30% speedup while retaining 90% accuracy	
YOLO-SED	Speed	FPS	YOLOv4	Use-case dataset	1 FPS on NVIDIA Jetson NX	iKPI-3-2 iKPI-3.3
RBAD	Speed	FPS	-	Use-case dataset	10 FPS	iKPI-3.2 iKPI-3.3

2.6.4.1. AVAD, ViAD, YOLO-SED and RBAD

The change for AVAD is due to the creation of the first publicly available audio-visual anomaly detection dataset by MARVEL[5], which was used for evaluations. The change for ViAD is due to the use of a more suitable dataset. RBAD receives object detection input from CATFlow and implements user-defined rules, thus the baselines are omitted. The new component YOLO-SED combines object detection and sound event detection functionalities and needs to be deployed at the edge.

2.6.5 Use Case KPIs

Table 17 reports the revised use case KPIs with respect to D1.2. The unmodified KPIs are not reported here.

Table 17: GRN R2 functional KPIs

Use case	KPI	Metric	Baseline	Expected result/ Improvement	Evaluators
GRN1	GRN-KPI1: Detect Vulnerable Road Users (VRUs) at any time of day, including	Detection rate/F1	Labelled dataset.	Improvement in detection rate due	GRN developers.

[5] MARVEL - Malta Audio Visual Anomaly Dataset (MAVAD), 2023. https://doi.org/10.5281/zenodo.7950008

	during low-light conditions			to multimodal detection	
	GRN-KPI2: low latency (time between detecting the cyclist and informing the road users)	Time in seconds	No baseline	2 seconds	GRN developers and managers
GRN2	GRN-KPI3: Automatically detect and label or quantify actions that determine driver behaviour	Human evaluation of output from models precision/recall	No baseline	4 improper behaviours detected	GRN developers and managers
UNS2	UNS-KPI5 Accuracy	Location-dependent error rate and F1-score	Baseline of the SELD task at DCASE Challenge	10% relative improvement on metrics in DCASE setup (dataset, cross-validation setup)	UNS staff
	UNS-KPI6 Accuracy	Classification-dependent localisation error and recall	Baseline of the SELD task at DCASE Challenge	10% relative improvement on metrics in DCASE setup (dataset, cross-validation setup)	UNS staff
	UNS-KPI7 Reaction time	Time	Human audio localisation and detection of the event	Decrease in time	UNS staff

The main changes to the functional KPIs of GRN1 were to include pedestrians as well, in addition to cyclists. This ensures that the use case has a greater impact on a wider range of vulnerable road users. The expected result is to note an improvement in the detection rate due to multimodal detection.

For GRN2 a more realistic and relevant expected result was set to detect 4 improper behaviours.

2.6.6 Use Case non-functional KPIs

The following KPIs have been revised with respect to D1.2 and D6.1.

2.6.6.1 GRN1: Safer Roads

Table 18 shows the new updated non-functional KPIs for GRN1.

Table 18: GRN1 non-functional KPIs

Use case	Evaluation variable	How to measure	Internal evaluators	External Evaluators
GRN1	End-user Experience Safer cycling and walking System welcomed by cyclists and pedestrians	Survey	GRN managers	Cyclists and pedestrians

	Efficacy	Car drivers welcome the system and agree that it will help them be more careful through a survey	GRN managers	Car drivers
	Scalability	Cost to add new devices/junctions	GRN managers	Third party service Providers

The main changes to the non-functional KPIs were to include pedestrians in addition to cyclists. This ensures that the use case has a greater impact on a wider range of vulnerable road users. In addition, the efficacy of the system is to be tested through a survey targeting car drivers.

2.6.6.2 GRN2: Road User Behaviour

Table 19 shows the new updated non-functional KPIs for GRN2.

Table 19: GRN2 non-functional KPIs

Use case	Evaluation variable	How to measure	Internal evaluators	External Evaluators
GRN2	Potential end-user	Survey	GRN Managers	Transport authorities Road users
	Efficacy	System evaluated by an NGO/public authority to determine the efficacy of the system when compared to what they use now to judge the success of a campaign	No baseline	Effectiveness verified by road experts in terms of the value it adds to a potential education campaign

The main change is evaluating the efficacy through a survey and understanding the added value to a potential educational campaign.

2.6.6.3 GRN3: Traffic Conditions and Anomalous Events

No changes were made to the Use case KPIs since D6.1

2.6.6.4 GRN4: Junction Traffic Trajectory Collection

No changes were made to the Use case KPIs since D6.1

2.6.6.5 MT1: Monitoring of Crowded Areas

No changes were made to the Use case KPIs since D6.1

2.6.6.6 MT2: Detecting Criminal and Anti-Social Behaviours

No changes were made to the Use case KPIs since D6.1

2.6.6.7 MT3: Monitoring of Parking Places

No changes were made to the Use case KPIs since D6.1

2.6.6.8 MT4: Analysis of a Specific Area

No changes were made to the Use case KPIs since D6.1

2.6.6.9 UNS1: Drone Experiment

No changes were made to the Use case KPIs since D6.1

2.6.6.10 UNS2: Localising Audio Events in Crowds

In comparison to the plans presented in D1.2 and D6.1, UNS2 use case was redefined as described in Section 2.2.3. Table 20 reports other non-functional KPIs for the redefined use case.

Table 20: UNS2 non-functional KPIs

Use case	Evaluation variable	How to measure	Internal evaluators	External Evaluators
UNS2	Modularity	Integration of new equipment	UNS staff	IT experts of security crews
	Data protection	Periodic evaluations	UNS staff	IT experts of security crews
	End-user experience	Periodic surveys	Infrastructure managers	Potentially public administration or public events organisers
	Scalability/Modularity	Extend the solution to the larger number of microphone boards placed in vicinity and exploit a smaller number of microphones per board	UNS staff	IT experts of security crews

3 GRN1: Safer Roads

This section describes the steps taken to integrate the components such that the GRN1: Safer Roads use case could be set up and tested. The following sections describe the framework configuration and data streams, the multimodal and privacy-aware intelligence applied, the demonstration of the implementation realised, and the work carried out.

3.1 Framework configuration and data streams for GRN1

3.1.1 Components

The implemented components in GRN1 are summarised in Table 21.

Table 21: MARVEL components in the GRN1

GRN1: Safer Roads			
Component owner	**Subsystem /Component**	**Comments on how the component is used in GRN1 for R2**	**Deployment location**
Sensing and perception subsystem			Edge and Fog
ITML	AV Registry	AV Registry contains metadata information of all AV sources present in GRN1. These include the information on the raw streams produced by the cameras (fps rate, resolution, etc.), but also on the anonymised streams produced by the VideoAnony anonymisation component	GRN Fog server (GRN F2)
GRN	Cameras with integrated microphones	One camera with integrated microphones is used in GRN1 to produce audio-visual streams from the Mgarr location	Edge
GRN	Jetson Edge Device	Apollo Dev kit to process audio video data at the edge and control the road sign	Edge
GRN	Arduino + LED	Arduino Nano running a script to control the LED array which simulates the road Traffic sign	Edge
GRN	Arduino Proxy	Implements an MQTT client for receiving messages from an MQTT broker and transforming them to a suitable format and transmitting over Serial protocol communication to an Arduino board	Edge
Security, privacy, and data protection subsystem			Edge, fog and cloud
FORTH	EdgeSec VPN	EdgeSec VPN creates a secure E2F2C VPN traffic backbone for all communications within the elements of the MARVEL platform by 100% encryption of the traffic	PC-simulated edge (GRN E1, E3), GRN Fog server (GRN F2), and Cloud (PSNC HPC via OpenStack)
Data management and distribution subsystem			Edge, Fog, and Cloud
ITML	Data Fusion Bus (DFB)	DFB stores inference results of the AI components that participate in GRN1	Cloud (PSNC HPC via OpenStack)
INTRA	StreamHandler	StreamHandler receives continuously anonymised AV data streams and segments them for temporary storage, for later on-request visual inspection through SmartViz.	GRN Fog server (GRN F2)

ATOS	DatAna Edge, Fog, and Cloud	DatAna for GRN1 consists of DatAna Edge, DatAna Fog, and DatAna Cloud components, each continuously consuming, through MQTT, the results of AI inference components deployed at the corresponding layer (Edge/Fog/Cloud) and sending to the relevant Kafka topics at the DFB	Jetson device (GRN E3), GRN Fog server (GRN F2), and Cloud (PSNC HPC)
Audio, visual, and multimodal AI subsystem			Edge, Fog, and Cloud
GRN	Arduino LED Proxy	A proxy to receive messages via MQTT and transmit them via a serial connection to the Arduino Nano which controls the LED	Edge (GRN E3)
AU	YOLO-SED	YOLOv4 component, which detects VRUs and SED to help distinguish between bicycles and motorcycles through audio. A logic component was also created to combine the multi-modal output	Edge (GRN E3)
Optimised E2F2C processing and deployment subsystem			Cloud
FORTH	MARVdash	MARVdash provides a Kubernetes-based deployment environment of all the GRN1 components. In GRN1, all nodes operate under Kubernetes/MARVdash	Cloud (PSNC HPC via OpenStack)
System outputs: User interactions and the decision-making toolkit			Cloud
ZELUS	SmartViz	SmartViz visualises detected anomalous road conditions which may be related (passively or actively) to obstructions	Cloud (PSNC HPC via OpenStack)

3.1.2 Pilot E2F2C infrastructure

The Jetson Apollo Dev Kit edge device is a development kit for the NVIDIA JetsonXavier NX used for GRN1. An Apollo Dev Kit of the same make was provided to AU to facilitate integration. Table 22 shows the specifications of the Jetson Device. This device is connected to an Arduino Nano, which controls a LED board that alerts vehicle drivers.

Table 22: GRN Edge Jetson device specifications

HW subsystem	Specifications
CPU	6-core NVIDIA Carmel ARM®v8.2 64-bit CPU 6MB L2 + 4MB L3 processor
GPU	NVIDIA Volta™ architecture with 384 NVIDIA CUDA® cores and 48 Tensor cores
Hard Drive	16 GB eMMC 5.1
RAM	8 GB 128-bit LPDDR4x @ 1600 MHs 51.2GB/s16 GB 128-bit LPDDR4x @ 59.7GB/s

All Edge devices are connected to 4G mobile routers.

GRN has set up a server which was provided by FORTH, as the GRN Fog node. This node processes anonymised streams for all the use cases, including GRN1. Table 23 shows the specifications of the GRN Fog server node.

Table 23: GRN Fog server node specifications

HW subsystem	Specifications
CPU	AMD EPYC 7313P 16-Core Processor
GPU	NVIDIA RTX A4000 16GB
Hard Drive	SSDs 3.5TB
RAM	252GB of RAM

The Fog server is connected to the internet via a wired connection.

3.1.3 Analysis of real-life data streams

IP Cameras - Edge

The GRN pilot includes three IP cameras. One camera is at Mgarr (a rural town in the north-west region), and the other two cameras are at Zejtun (an urban town close to the southern inner-harbour region). These cameras have no ability to process data at the edge and can only transmit audio and video via an IP connection. The camera at Mgarr has slightly different specifications than the IP cameras at Zejtun. All three components will be used in both GRN use cases employed for R2, namely GRN1 and GRN2.

Table 24 lists the specifications for each camera.

Table 24: Specifications for GRN IP Cameras

Specifications Type	Mgarr Camera	Zejtun Cameras
IP camera model	Safire 5MP Bullet Outdoor/Indoor IP Camera With PoE	ANNKE outdoor 5MP PoE security cameras, Model I51DL, Lens: 2.8mm
Resolution	1920 x 1080 P	1920 x 1080 P
Frame Rate	25 fps	20 fps
Video Encoding	H.264	H.264
Audio Encoding	MP2L2	MP2L2
Audio Sampling Rate	32kHz	16kHz
Audio Stream Bitrate	64kbps	128kbps

Only the Mgarr cameras were used for this use case, however, the entire camera specifications were listed for completeness.

3.2 Multimodal and privacy-aware intelligence for GRN1

3.2.1 Datasets for model training and privacy assurance

This section describes the methods and approaches taken towards collecting the AV data, the datasets required for model training, the analysis of datasets and streams, and the privacy assurance and anonymisation methods used in the use case.

3.2.1.1 Datasets for model training

GRN has provided data in various forms to support the development of this use case.

For the YOLO component in the YOLO-SED model, GRN has provided the pre-trained weights from CATFlow to avoid retraining.

The dataset for SED training is the GRN-AV-traffic-entity dataset described in D2.1[6]. For this dataset, audio-visual snippets were obtained from various locations around Malta and from the GRN static cameras. More details on the data labelling scheme can be found in D6.1, Section 1.1.1.4.2.1.1. For the R2 integration, efforts were made to increase the number of examples for sparse classes such as motorcycles and buses.

A small-scale multimodal dataset is provided for testing the YOLO-SED component. This dataset contains examples of pedestrians, bicycles, and motorcycles labelled with bounding boxes and sound events.

3.2.1.2 Analysis of datasets

The analysis of the streams and datasets is performed by the AI model providers. Pending feedback, more data for training will be added.

3.2.1.3 Privacy assurance and anonymisation

To ensure privacy, all video streamers are anonymised using VideoAnony. The current version of this component blurs number plates and faces resulting in AV streams that contain no identifiable information.

3.3 Demonstration

3.3.1 The Decision-making Toolkit

GRN1 is a use case that focuses on enhancing the safety of vulnerable road users, specifically cyclists and pedestrians in Malta, with the objective of promoting active travel modes. To achieve this goal, a low-latency road traffic sign is utilised, which illuminates to alert drivers whenever a vulnerable road user is detected. The detections from YOLO-SED are transmitted to the MARVEL platform for visualisation in SmartViz, enabling remote monitoring of the situation.

The dashboard of the DMT, as shown in Figure 2, is designed to support this use case. Real-time alerts, detected by YOLO-SED, are sent to SmartViz through Kafka messages via the Data Fusion Bus (DFB) component. These alerts are then visualised in the Alerts widget, providing users with immediate information to facilitate prompt actions. Additionally, the Details widget displays the last detected events in a textual format, utilising historical data accessed from an Elastic search database. This database is continuously updated by the audio, visual, and multimodal AI subsystem components.

[6] "D2.1: Collection and Analysis of Experimental Data," Project MARVEL, 2021. https://doi.org/10.5281/zenodo.5052713

Within SmartViz, users have the capability to select an event displayed in the widget and play the corresponding video snippet. The StreamHandler component facilitates this functionality by retrieving the relevant video from the camera stream within the specific time period of the detection. The segmented stream generates a URL containing the video for the selected event, which is played in SmartViz using the "Anomaly and Event Detection player" widget.

During video playback, users have the ability to validate the inference result by marking them as accurate or not. This validation process contributes to the continuous improvement of the system and the training of AI models. The inference result verification is sent through Kafka messages from SmartViz to DFB, which updates the status of the corresponding event and stores the information in an Elastic search index accessed by the Data Corpus.

In addition to widget arrangement and resizing, the dashboard view offers functionality for users to download the visualised data in JSON format. This feature allows for easy retrieval and utilisation of the data for further analysis and reference purposes. Furthermore, users have the option to save the entire dashboard as a PDF file, enabling offline access and sharing based on their individual preferences.

Finally, the Weather Information widget in this use case provides users with a representation of weather-related data. It allows users to view and explore weather information for a selected time period. By visualising weather variables such as visibility, humidity, temperature, and overall weather conditions, users can gain insights and uncover potential correlations between detections of events and anomalies and weather data that may influence them.

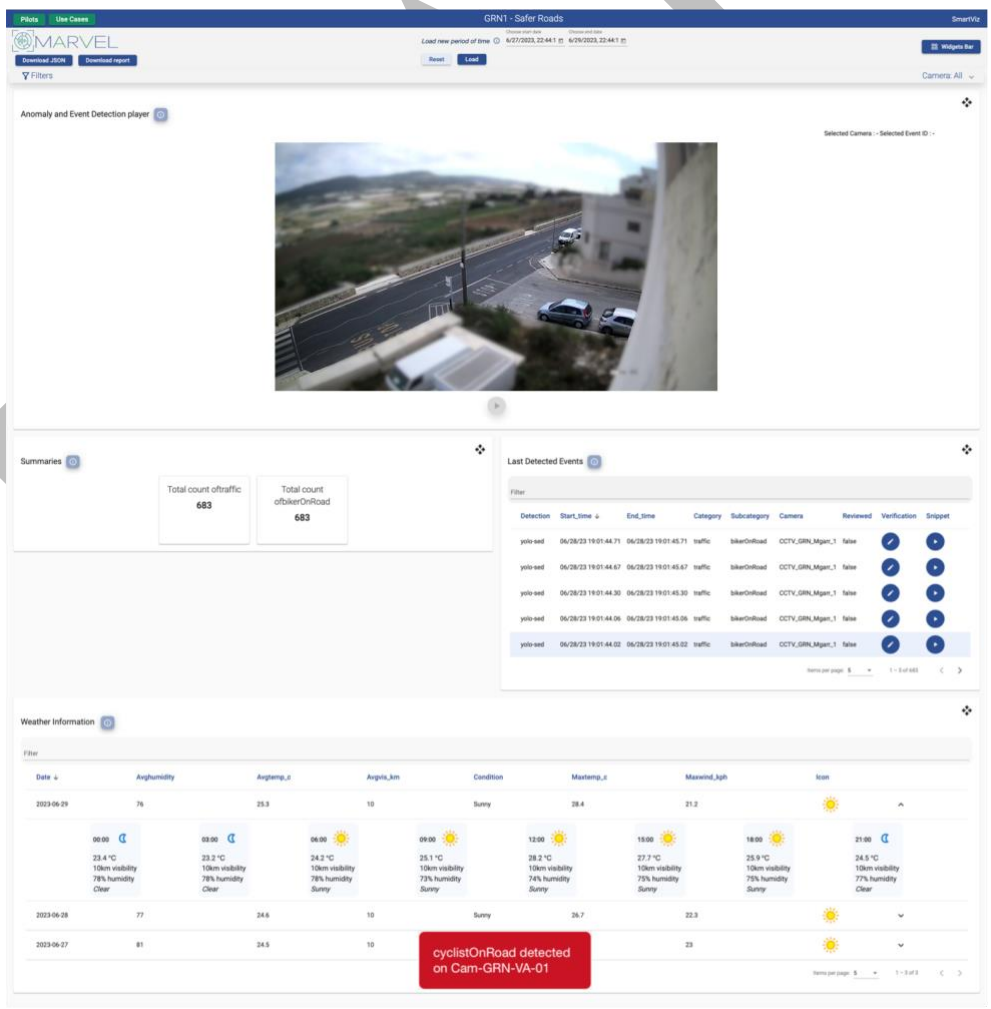

Figure 2. GRN1 in the DMT

D4.6[7] reports the complete functionalities, widgets, and users requirements regarding DMT for GRN1.

3.4 Work carried out

GRN has been actively organising work to see to the completion of the GRN use cases. These efforts included the creation of a detailed spreadsheet to collect all the information necessary for component owners and technical partners per use case, the lead of the GRN pilot-focused meetings to discuss relevant topics in detail, and active attendance in the technical and integration meetings.

All information on the GRN use cases was collected in an interactive Google Sheet that was shared with partners. This format allowed GRN leaders to constantly update the sheet with new information and allowed other partners to comment and ask questions. Separate sheets were created for each identified issue: action items, GRN infrastructure, details on each of the use cases, AI models required, diagrams depicting the proposed dataflow, information about the relevant events that should be detected, labelled datasets available and in progress, user stories in detail, and the demonstrators per use case.

Until now, GRN has chaired eight pilot-focused meetings for the R2 interaction efforts. Each meeting started with addressing all action items that were pending. The meetings were useful to (a) re-introduce the two new use cases for the R2 integration (goals, data requirements, and the AI models required); (b) discuss if the partners had the necessary hardware for development; (c) present the GRN infrastructure, use case data flows, data requirements, and the user interface; (d) update wrap up the R1 use cases. Occasionally, the pilot-focused meeting was repurposed for technical or integration meetings if no GRN use case needed further discussion.

The GRN partners were also active in technical meetings. GRN often plans and leads meetings discussing the UI and specific AI components. GRN was also an active participant in all integration meetings. Apart from being the TAD and CATFlow owner, GRN also had to constantly monitor and support the infrastructure to make sure all meetings could go ahead as planned.

[7] "D4.6 - MARVEL's decision-making toolkit – final version," Project MARVEL, 2023. https://doi.org/10.5281/zenodo.8147077

4 GRN2: Road User Behaviour

This section describes the steps taken to integrate the components such that the GRN2: Road User Behaviour use case could be set up and tested. The following sections describe the framework configuration and data streams, the multimodal and privacy-aware intelligence applied, the demonstration of the implementation realised, and the work carried out.

4.1 Framework configuration and data streams for GRN2

4.1.1 Components

The implemented components in GRN2 are summarised in Table 25.

Table 25: MARVEL components in the GRN2

GRN2: Road Users Behaviour			
Component owner	**Subsystem /Component**	**Comments on how the component is used in GRN2 for R2**	**Deployment location**
Sensing and perception subsystem			Edge and fog
ITML	AV Registry	AV Registry contains metadata information of all AV sources present in GRN2. These include the information on the raw streams produced by the cameras (fps rate, resolution, etc.), but also on the anonymised streams produced by the VideoAnony anonymisation component	GRN Fog server (GRN F2)
GRN	Cameras with integrated microphones	Three cameras with integrated microphones are used in GRN2 to produce audio-visual streams. Two cameras are used at the Zejtun and one at the Mgarr location	Edge cameras
Security, privacy, and data protection subsystem			Edge, Fog, and Cloud
FORTH	EdgeSec VPN	EdgeSec VPN creates a secure E2F2C VPN traffic backbone for all communications within the elements of the MARVEL platform by 100% encryption of the traffic	PC-simulated Edge (GRN E1, E2), GRN Fog server (GRN F2), and Cloud (PSNC HPC via OpenStack)
FBK	VideoAnony	VideoAnony detects the cyclists and pedestrians that appear in the scene and anonymises their faces. In GRN2, the component is present with three instances, once for each of the three camera streams	PC-simulated Edge (GRN E1, E2) and GRN Fog server (GRN F2)
Data management and distribution subsystem			Edge, Fog, and Cloud
ITML	Data Fusion Bus (DFB)	DFB stores inference results of the AI components that participate in GRN2	Cloud (PSNC HPC via OpenStack)
INTRA	StreamHandler	StreamHandler receives continuously anonymised AV data streams and segments them for temporary storage, for later on-request visual inspection through SmartViz	GRN Fog server (GRN F2)
ATOS	DatAna Edge, Fog, and Cloud	DatAna for GRN2 consists of DatAna Edge, DatAna Fog, and DatAna Cloud components, each continuously consuming, through MQTT, the results of AI inference components deployed	PC-simulated Edge (GRN E1, E2), GRN Fog server (GRN

		at the corresponding layer (Edge/Fog/Cloud) and sending to the relevant Kafka topics at the DFB	F2), and Cloud (PSNC HPC)
CNR	HDD	For the predefined GRN2 AI inference components, HDD optimises the allocation of their output streams across a given set of DFB Kafka topics	Cloud (PSNC HPC via OpenStack)
Audio, visual, and multimodal AI subsystem			Edge, Fog, and Cloud
GRN	CATFlow+TAD	CATFlow classifies traffic entities (different vehicle classes and pedestrians) and provides their trajectories. TAD detects traffic anomalies, specifically, high and low traffic speeds, after processing CATFlow outputs	PC-simulated Edge (GRN E1, E2) and GRN Fog server (GRN F2)
AU	Rule Based anomaly detector (RBAD)	In GRN2, RBAD component detects anomalies through the application of a set of predefined rules on the CATFlow output	PC-simulated Edge (GRN E1, E2) and GRN Fog server (GRN F2)
TAU	Sound Event Detection (SED)	In GRN2, SED detects vehicle horns	PC-simulated Edge (GRN E1, E2) and GRN Fog server (GRN F2)
Optimised E2F2C processing and deployment subsystem			Cloud
FORTH	MARVdash	MARVdash provides a Kubernetes-based deployment environment of all the GRN2 components. In GRN2, all nodes operate under Kubernetes/MARVdash	Cloud (PSNC HPC via OpenStack)
System outputs: User interactions and the decision-making toolkit			Cloud
ZELUS	SmartViz	SmartViz visualises detected anomalous road conditions which may be related (passively or actively) to obstructions	Cloud (PSNC HPC via OpenStack)

4.1.2 Pilot E2F2C infrastructure

The GRN Infrastructure provided for the R2 integration consists of three IP cameras transmitting audio and video, two PCs to simulate the Edge layer computational, nodes and a server provided by FORTH acting as the Fog computational node. The following paragraphs describe each component in detail.

GRN has a PC at the Mgarr location directly connected to the Mgarr IP Camera. Table 26 shows the specifications of the GRN Edge PC 1. This PC was used to carry out processing at the Edge.

Table 26: GRN Edge PC 1 specifications

HW subsystem	Specifications
CPU	Intel Core i7-3770
GPU	GTX1650
Hard Drive	1 TB
RAM	32 GB

A second PC was added to the Zejtun location to directly connect to the Zejtun IP Cameras. Table 27 shows the specifications of the GRN Edge PC 2. This PC was used to carry out processing at the edge and further avoid the transfer of anonymised data.

Table 27: GRN Edge PC 2 specifications

HW subsystem	Specifications
CPU	Intel Core i7 12700
GPU	Geforce RTX 3080
Hard Drive	1 TB
RAM	16GB

All Edge devices are connected to 4 G mobile routers.

GRN has set up a server, provided by FORTH, as the GRN Fog node. This node processes anonymised streams for all the use cases. Specifications of the GRN Fog server node can be found in Table 23 in Section 3.1.2.

4.1.3 Analysis of real-life data streams

Details about the GRN real-life data streaming system can be found in Section 3.1.3.

4.2 Multimodal and privacy-aware intelligence for GRN2

4.2.1 Datasets for model training and privacy assurance

This section describes the methods and approaches taken towards collecting the AV data, the datasets required for model training, the analysis of datasets and streams and privacy assurance and anonymisation methods used in the use case.

4.2.1.1 Datasets for model training

GRN provided a dataset for the testing of the TAD component. This dataset was collected from the Mgarr location. The annotators were instructed to go through a series of five-minute videos and find cases where vehicles underwent unusual trajectories, given the context. This could include examples of a vehicle making a U-turn, parking, and just taking too long to get out of the stop sign. The annotators also labelled examples of rare vehicles, such as bicycles and scooters. Any time an anomaly was detected, the name of the video and anomaly start and end time stamps were recorded in a spreadsheet. This dataset differs from the AVAD dataset since the anomalies were not cropped out from the video, as cropping them out from the video results in parts of the trajectories being missing. The datasets consist of 137 examples of labelled anomalous events which amount to more than 11 hours of data.

In addition, GRN provided a small-scale dataset for testing the SED horn detection model. This dataset consists of 13 examples of horn sounds annotated using ELAN software.

4.2.1.2 Analysis of datasets

The analysis of the streams and datasets is performed by the AI model providers. Pending feedback, more data for training will be added.

4.2.1.3 Privacy assurance and anonymisation

To ensure privacy, all video streams are anonymised using VideoAnony. The current version of this component blurs number plates and faces thus the streams contain no identifiable information.

4.3 Demonstration

4.3.1 The Decision-making Toolkit

GRN2 focuses on monitoring the behaviour of road users at a junction to support law enforcement and educational campaigns promoting responsible driving, cycling, and other road uses. The use of AI models through the MARVEL platform enables automatic detection of common bad behaviours, providing accurate counts and comparisons of occurrences over time.

The SmartViz platform, shown in Figure 3, displays the detected behaviours and allows users to select two time periods for comparison. The Summaries widget presents the total number of detected anomalies, utilising outputs from SED, RBAD, and only the anomalous events of TAD. The Statistics widget visualises bar charts indicating the total number of detections per anomalous event.

The Temporal Representation widget presents the information of detections by all AI components used in this use case in a temporal format. The Details widget visualises the detections in a textual format, with CATflow and TAD displayed in the "Traffic Events Detection" table, and RBAD and SED displayed in the "Sound Events and Anomalies Detection" table.

All data feeding the widgets in this use case are historical and accessed through an Elastic search database, which is constantly updated by the audio, visual, and multimodal AI subsystem components.

Within the audio and video player functionality in SmartViz, users can select an event and play the corresponding snippet. The StreamHandler component facilitates this feature by retrieving the stream for the relevant time period of the detection. The segmented stream generates a URL containing the snippet for the selected event, which is played in the Audio or Video player widget within SmartViz.

Users have the ability to validate the inference result by marking them as accurate or not. The inference result verification is sent through Kafka messages from SmartViz to the Data Fusion Bus (DFB), which then updates the status of the corresponding event and stores the information in an Elastic search index accessed by the Data Corpus.

In addition to rearranging and resizing widgets, the dashboard view offers functionality for users to download the visualised data in JSON format. Furthermore, users can save the entire dashboard as a PDF file, allowing for offline access and sharing according to their preferences.

Finally, the Weather Information widget in this use case provides users with a representation of weather-related data, allowing them to view and explore weather information for a selected time period.

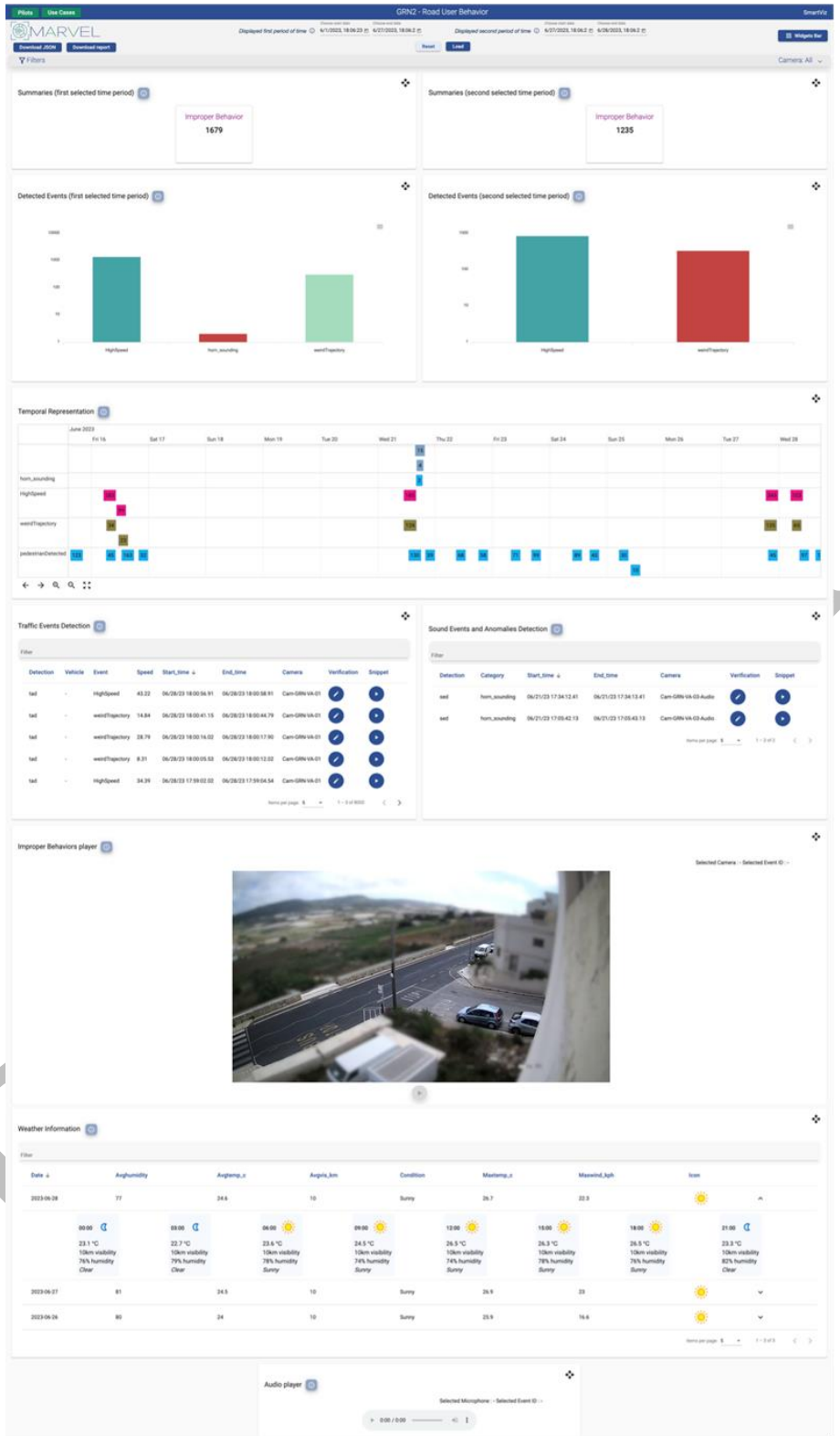

Figure 3. GRN2 in the DMT

D4.6 contains the complete functionalities, widgets, and users requirements regarding DMT for GRN2.

4.4 Work carried out

For a detailed description of the work done, please refer to Section 3.4 of this document.

5 GRN3: Traffic Conditions and Anomalous Events

GRN3 use case is briefly described in Section 2.3.1 of this document.

5.1 Updates compared to R1

GRN3 underwent two significant updates to enhance its functionality. The first focused on improving the user interface after providing feedback to the ZELUS team. The objective was to present the data in a more user friendly manner. Please refer to D4.6 to see the complete functionalities, widgets and users requirements regarding Decision-making Toolkit.

The second update involved infrastructure changes. To ensure efficient anonymisation of data near the cameras and to augment computational resources, an Edge Device was added in the Zejtun location. The GRN Fog was also upgraded to a more computationally powerful server provided by FORTH, thereby enhancing overall computational capabilities. All the infrastructure changes were tested.

Furthermore, the AT component was transferred from the Cloud layer to the Fog layer.

One instance of CATFlow and TAD, responsible for processing one of the two Zejtun streams, was also transferred from the Fog layer to the Edge layer. The new Edge Device in Zejtun also hosts a dedicated DatAna agent.

6 GRN4: Junction Traffic Trajectory Collection

GRN4 use case is briefly described in Section 2.3.2 of this document.

6.1 Updates compared to R1

GRN4 underwent updates similar to GRN 3, which involved improvements to the user interface and corresponding infrastructure enhancements. To gain comprehensive insights into the Decision-making Toolkit's complete functionalities, widgets, and user requirements, please refer to D4.6, where detailed information is provided.

One instance of CATFlow and TAD processing one of the two Zejtun streams was also transferred from the fog layer to the edge layer. The new Edge Device in Zejtun also hosts a dedicated DatAna agent.

7 MT1: Monitoring of Crowded Areas

MT1 use case is briefly described in Section 2.3.3 of this document.

7.1 Updates compared to R1

The MT1 use case underwent updates by providing valuable feedback to the ZELUS team, leading to improvements in the user interface and the presentation of data in a more user-friendly manner. For a comprehensive understanding of the complete functionalities, widgets, and user requirements of the Decision-making Toolkit, please refer to D4.6, where detailed information is available.

8 MT2: Detecting Criminal and Anti-Social Behaviours

This section describes the steps taken to integrate the components such that the MT2: Detecting Criminal and Anti-Social Behaviours use case could be set up and tested. The following sections describe the framework configuration and data streams, the multimodal and privacy-aware intelligence applied, the demonstration of the implementation realised, and the work carried out.

8.1 Framework configuration and data streams for MT2

8.1.1 Components

Table 28 lists the components deployed in the use case MT2. The table briefly describes the functionalities of the component that are relevant to this particular use case. A detailed description of the architecture of the use case is available in D5.6, while the operational and implementation details of the components are reported in the related technical and scientific deliverables.

Table 28: MARVEL components in the MT2

MT2: Detecting Criminal and Anti-Social Behaviours			
Component owner	**Subsystem /Component**	**Comments on how the component is used in MT2 for R2**	**Deployment location**
Sensing and perception subsystem			Edge and Fog
ITML	AV Registry	AV Registry contains metadata information of all AV sources present in MT2. These include the information on the raw streams produced by the camera and the microphone (fps rate, resolution, etc.), but also on the anonymised streams produced by the VideoAnony and AudioAnony anonymisation components.	FBK Fog WS PC (MT F2-Kubernetes)
IFAG	MEMS microphone	Two MEMS microphones Nano Hub connected to a Raspberry Pi, are used in MT2, specifically in Piazza Santa Maria Maggiore.	Edge microphone outside Kubernetes
MT	Camera	Two cameras with enabled streaming are used in MT2, specifically in Piazza Santa Maria Maggiore.	Edge camera
Security, privacy, and data protection subsystem			Edge, Fog, and Cloud
FORTH	EdgeSec VPN	In MT2, EdgeSec VPN creates a secure F2C VPN traffic backbone for all communications within the elements of the MARVEL platform by 100% encryption of the traffic. For MT2, the edge layer is not part of the EdgeSec VPN network.	FBK Fog WS PC (MT F2-Kubernetes) and Cloud (PSNC HPC via OpenStack)
FBK	AudioAnony+VAD	Based on the onset and offset times of speech segments detected by VAD, AudioAnony will anonymise the respective segments of the audio. The component runs on the Raspberry Pi, outside Kubernetes. The audio signals are only available after anonymisation. As MT2 uses one audio stream, the component is present with one instance.	Raspberry Pi (MT2 E1 and MT2 E2)

Data management and distribution subsystem			Edge, Fog, and Cloud
ITML	Data Fusion Bus (DFB)	DFB stores inference results of the AI components that participate in MT2.	Cloud (PSNC HPC via OpenStack)
INTRA	StreamHandler	StreamHandler receives continuously anonymised AV data streams and segments them for temporary storage, for later, on-request visual inspection through SmartViz.	FBK Fog WS PC (MT F2-Kubernetes)
ATOS	DatAna Fog and Cloud, MQTT Edge	DatAna for MT2 consists of DatAna Fog and DatAna Cloud components, each continuously consuming, through MQTT, the results of AI inference components deployed at the corresponding layer (Edge/Fog/Cloud) and sending to the relevant Kafka topics at the DFB	Raspberry Pi (MT E1), FBK Fog WS PC (MT F2-Kubernetes) and Cloud (PSNC HPC via OpenStack)
CNR	HDD	For the predefined MT2 AI inference components, HDD optimises the allocation of their output streams across a given set of DFB Kafka topics.	Cloud (PSNC HPC via OpenStack)
Audio, visual, and multimodal AI subsystem			Cloud
AU	Audio-Visual Anomaly Detection (AVAD)	In MT2, AVAD detects anomalies in the square (e.g., drug dealing, people fighting)	Cloud (PSNC HPC via OpenStack)
TAU	Automated Audio Captioning (AAC)	In MT2, AAC detects and describes audio events that occur in the place (e.g., people arguing, people shouting, etc.)	FBK Fog WS PC (MT F2-Kubernetes)
TAU	Sound Event Detection (SED)	In MT2, SED detects and describes audio events that occur in the place (e.g., people fighting, people shouting, etc.)	Cloud (PSNC HPC via OpenStack)
TAU	Audio Tagging (AT)	In MT2, AT outputs audio tags for the acoustic environment of the monitored locations	Cloud (PSNC HPC via OpenStack)
FORTH	GPURegex	In MT2, GPURegex searches for specific patterns against captioning data resulting from AAC. If there is a match, GPURegex reports an alert. Each alert indicates a criminal or anti-social event	FBK Fog WS PC (MT F2-Kubernetes)
Optimised E2F2C processing and deployment subsystem			Cloud
FORTH	MARVdash	MARVdash provides a Kubernetes-based deployment environment for all MT2 components operating under part of the MT infrastructure managed by Kubernetes	Cloud (PSNC HPC via OpenStack)
System outputs: User interactions and the decision-making toolkit			Cloud
ZELUS	SmartViz	SmartViz is a UI/UX which presents the anomalous events detected by the AI components in Piazza Santa Maria Maggiore in a user-friendly manner	Cloud (PSNC HPC via OpenStack)

8.1.2 Pilot E2F2C infrastructure

The MT2 infrastructure provided for the R2 integration consists of two IP cameras transmitting video, two Raspberry Pis devices transmitting audio data collected from the connected IFAG microphones and two workstations at the fog layer managed by FBK.

The following subsections will describe each component in detail.

Edge devices

The MT2 edge layer infrastructure consists of two IP cameras transmitting video and two Raspberry Pi with a connected IFAG microphone. The Raspberry Pi runs a voice activity detector and a speech anonymiser and streams the anonymised audio via RTSP. The upload function is a secure transmission by VPN access between MT and FBK in which raw data will be sent to the data lake in FBK.

IFAG has provided two versions of the integrated microphone boards: the AudioHub Nano and the Audiohub - Nano 4 Mic Version to collect the audio data. The integrated microphone board stream mono, stereo or 4 channels audio data. Both integrated microphone boards use the standard USB Audio protocol, which is supported by the edge devices used in this demonstrator (Intel NUCs and Raspberry Pi). No driver installation is required, and recording can be performed by using common audio recording software and libraries. The user can choose the desired number of channels to record and the sampling rate. With the on-board switch, the operating mode and gain configuration can be selected (from 0 to 24dB) to better suit the recording scenario. An extended technical explanation of the devices can be found in deliverable D4.1[8], Section 2.2. The edge devices in MT2 employ the 2-microphone version.

Table 29 lists the specifications for each device.

Table 29: MT2 Sensing devices

Specifications Type	Piazza Santa Maria Maggiore
Video	
IP camera model	Digital cameras - Basler BIP-1600dn
Resolution	1600 x 1200
Frame Rate	12,5 fps
Video Encoding	H.264
Audio	
Microphones	IFAG-MEMS
Audio Encoding	ACC (LC)
Audio Sampling Rate	16kHz
Audio Stream Bitrate	Mono 69 kbps
RPi	Raspberry Pi 4 Model B 8GB RAM – Micro SD 32GB

[8] "D4.1: Optimal audio-visual capturing, analysis and voice anonymisation – initial version," Project MARVEL, 2020. https://doi.org/10.5281/zenodo.5833277

Fog workstations

FBK provides the Fog tier for the MT use cases. To comply with the constraints of the agreement between MT and FBK (FBK was nomineed as data processor) 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.

Table 30 lists the specifications of workstation 2. Workstation 1 is an FBK internal machine which may change and will not be accessed by users external to the research organisation.

Table 30: Specification of the FBK Fog workstation 2 in the Kubernetes cluster

Component	Specifications
CPU	Intel Xeon E5-1620
GPU	Tesla K40 11GB
Hard Drive	512GB
RAM	20GB

8.1.3 Analysis of real-life data streams

In the case of MT2, video data is recorded using two IP cameras of resolution 1600x1200 pixels. Data capturing is performed using 12,5 frames per second, H.264 codec is used, and data is stored using the MP4 format.

Audio is captured using IFAG-MEMS at a sampling frequency equal to 16kHz, whereas bit depth is 16 bits per sample. Taking into account that mono audio is recorded, the bit-rate is 69 kbps. Audio data is recorded using WAV format.

8.2 Multimodal and privacy-aware intelligence for MT2

This section describes the methods and approaches taken to collect AV data, the datasets required for model training, the analysis of datasets the privacy assurance and the anonymisation method used in the use case.

8.2.1 Datasets for model training and privacy assurance

8.2.1.1 Datasets for model training

MT has provided data for various components. MT has provided the dataset "TrentoOutdoor – real recording" and "TrentoOutdoor – staged recording" (as defined in D2.1) processing the streams of cameras and microphones from Piazza Santa Maria Maggiore.

MT, in collaboration with FBK, collected, anonymised, and annotated more than 180 audio and video files, in total 30 GB used for the training of:

- AAC and GPURegex – MT contributed 130 raw audio samples, which correspond to a total of 3 hours. The size of the samples is 11 GB in total. The samples contain both anomalous and non-anomalous situations. The files were annotated by 3 different annotators;
- SED and AT – MT annotated the same 130 raw audio samples used for AAC and GPURegex, but using the ontology needed to train SED and AT;
- AVAD – MT annotated 20 GB of video streams which contain anomaly and not anomaly situations.

Thanks to the staged recording done during M16, M28, and M29, MT and FBK have provided 40 audio-videos each manually annotated and used for the training of AAC, GPURegex, SED, AT, and AVAD.

8.2.1.2 Analysis of datasets

The analysis of the streams and datasets is performed by the AI model providers. Pending feedback, more data for training will be added.

8.2.1.3 Privacy assurance and anonymisation

The collected video data are anonymised at FBK premises using the VideoAnony tool. Instead, the collected audio data are anonymised by the composite component AudioAnony+VAD on edge devices (Raspberry Pis). In this way, MT and FBK comply with the privacy constraints defined in the appointment of FBK as data processor, and they can manage and share data with consortium partners.

For the staged recordings, anonymisation is not necessary as all involved subjects signed the informed consent.

8.3 Demonstration

8.3.1 The Decision-making Toolkit

This use case focuses on detecting criminal and anti-social behaviours and identifying dangerous situations such as robberies and aggressions. The integration of AI models through the MARVEL platform enables automatic detection of these behaviours, and SmartViz displays the detections in a dedicated dashboard (Figure 4).

To address the user's requirement of receiving alerts when anomalies occur, the Alerts widget has been integrated into this use case. Real-time alerts are detected by GPURegex and transmitted to SmartViz through Kafka messages via the DFB component. The Alerts widget visualises these alerts, allowing users to promptly respond and take appropriate actions based on the detected anomalies.

The Summaries widget provides an overview of the total number of detected anomalies and events, utilising outputs from SED, GPURegex, and AVAD components.

Specifically developed for the AAC component, the Word Cloud widget enables users to visualise the most common keywords and descriptions associated with the detected events.

Within the audio and video player functionality in SmartViz, users can select an event and play the corresponding audio or video snippet. The StreamHandler component retrieves the stream for the relevant time period of the detection, segments it, and plays the snippet in the Audio or Video player widget within SmartViz.

The Details widget presents the detections in a textual format, with SED and AT displayed in the "Sound Events and Anomalies Detection" table, and AAC displayed in the "Anomalous Audio Captioning" table.

Users have the ability to validate the inference results by marking them as accurate or not. The verification of inference results is sent through Kafka messages from SmartViz to the DFB, which then updates the status of the corresponding event and stores them in an Elasticsearch index accessed by the Data Corpus. The use case also includes a police intervention functionality, where users can mark an event as important and in need of police intervention, facilitating appropriate action.

In addition to rearranging and resizing widgets, the dashboard view offers the functionality to download the visualised data in JSON format. Furthermore, users have the option to save the entire dashboard as a PDF file, allowing for offline access and convenient sharing based on individual preferences.

Finally, the Weather Information widget provides users with a representation of weather-related data, enabling them to view and explore weather information for a selected time period, which may help uncover correlations between detected events and anomalies and weather conditions.

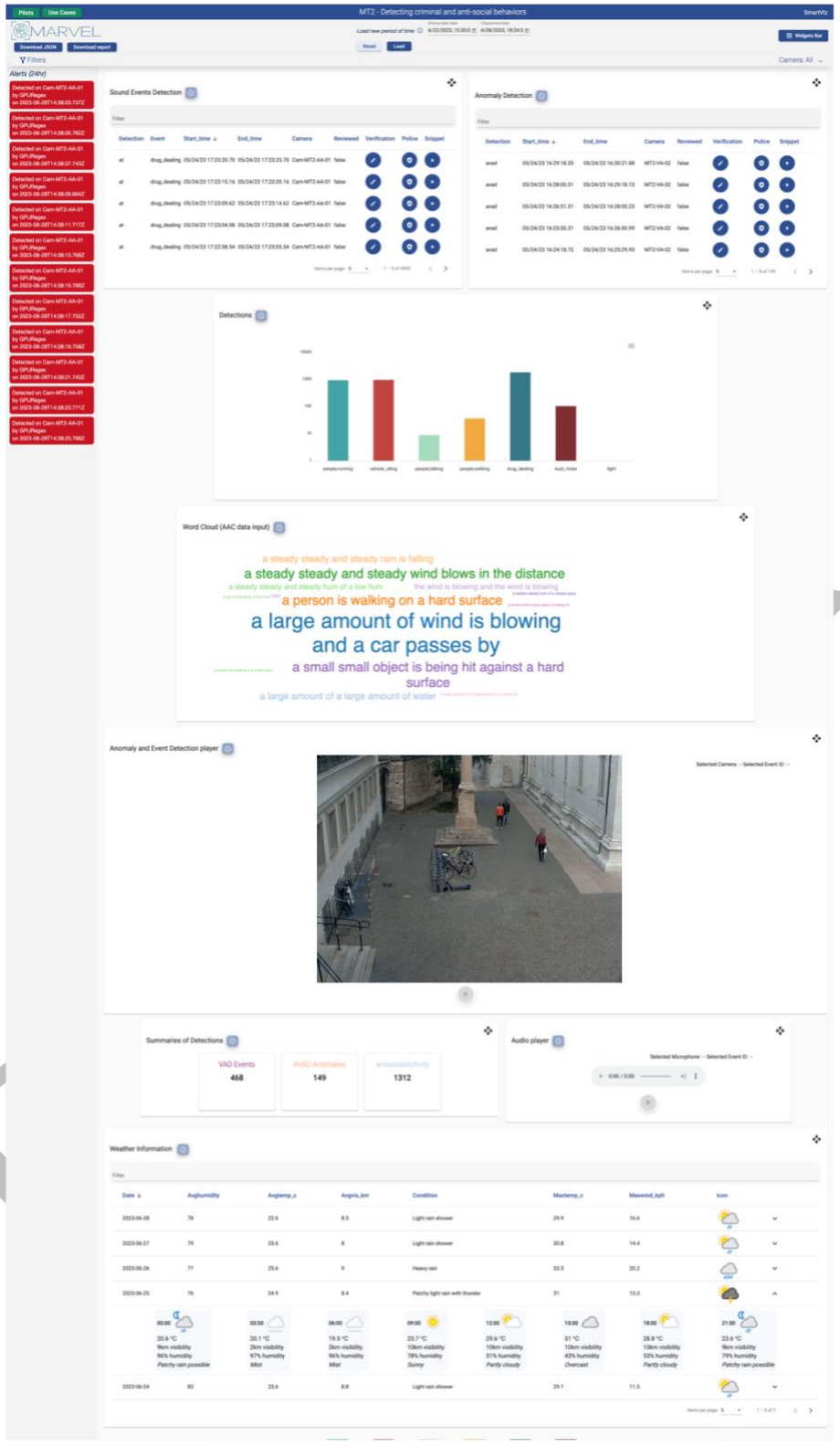

Figure 4. MT2 in the DMT

Please refer to D4.6 to see the complete functionalities, widgets, and users requirements regarding DMT for MT2.

8.4 Work carried out

MT has actively coordinated efforts to ensure the completion of the MT use cases. These efforts encompassed several activities, such as creating a comprehensive spreadsheet to gather all the necessary information for component owners and technical partners per use case. In addition, MT conducted pilot-focused meetings to delve into relevant topics, actively participated in technical and integration meetings, and diligently attended to the organisation of work.

All the information regarding the MT use cases was compiled in an interactive Google Sheet, which was shared with partners. This format allowed MT leaders to consistently update the sheet with new information, while also enabling other partners to provide comments and ask questions. Separate sheets were dedicated to specific aspects of the use cases, such as action items, MT infrastructure, detailed information on each use case, required AI models, diagrams depicting the proposed data flow, information about relevant events, available and ongoing labelled datasets, detailed user stories, and the respective demonstrators for each use case.

Until the time of publication, MT successfully chaired nine pilot-focused meetings focused on the R2 integration efforts. Each meeting commenced by addressing pending action items. These gatherings proved valuable in several ways, including: (a) introducing two new use cases for R2 integration, outlining their goals, data requirements, and the necessary AI models; (b) discussing whether partners possessed the requisite hardware for development; (c) presenting the MT infrastructure, use case data flows, data requirements, and user interface; (d) wrapping up the R1 use cases.

Furthermore, MT actively participated in all integration meetings, diligently monitoring and supporting the infrastructure to ensure the smooth progress of all scheduled meetings.

9 MT3: Monitoring of Parking Places

MT3 use case is briefly described in Section 2.3.4 of this document.

9.1 Updates compared to R1

The MT3 use case underwent updates by providing valuable feedback to the ZELUS team, resulting in improvements to the user interface and the presentation of data in a more user-friendly manner.

The required functionalities reported in D1.2 – Section 3.4.4 and 3.8.3 as well as in D6.1 – Section 2.3.4 and 7.4.1 were explored. It was planned that the "*anomalous behaviours will be examined such as, for example, the correct use of parking spaces reserved for taxis, the occupation of spaces reserved for the vehicles of disabled people, the number of parked campers, and their time of stay (also in relation to events such as Christmas markets and other public events planned in the city that attract tourists to enter the city), the average parking time of vehicles, the use of the cycle boxes installed in the area*" and also "*check if only taxis park in the taxi rank, check the disabled parking spaces, check the number of campers, the average length of stay*".

However, during the development and improvement of the use case, it was noted that these functionalities could not be implemented due to the angle of the available cameras (not enough parking space can be covered to carry out such analyses).

The implementation of parking monitoring, which was realised for R1, is still valid regarding the temporal distribution of vehicles in the car park, the total number of vehicles, the grouping of vehicles and/or events, and the information on observed detections. Please refer to D4.6 to see the full functionality, widgets and user requirements for the Decision-making Toolkit of this use case.

Finally, the MT3 infrastructure was also updated. This included the addition of one Jetson Apollo Dev Kit (refer to Table 22 for technical information) at the edge tier for more computing capacity for video streams. All the infrastructure changes were tested extensively to ensure the infrastructure is as stable as possible.

10 MT4: Analysis of a Specific Area

This section describes the steps taken to integrate the components such that the MT4: Analysis of a Specific Area use case could be set up and tested. The following sections describe the framework configuration and data streams, the multimodal and privacy-aware intelligence applied, the demonstration of the implementation realised and, the work carried out.

10.1 Framework configuration and data streams for MT4

10.1.1 Components

Table 31 lists the components deployed in the use case MT2. The table briefly describes the functionalities of each component that is relevant to this particular use case. The detailed description of the architecture of the use case is available in D5.6, while the operational and implementation details of the components are reported in the related technical and scientific deliverables.

Table 31: MARVEL components in the MT4

MT4: Analysis of a Specific Area			
Component owner	**Subsystem /Component**	**Comments on how the component is used in MT4 for R2**	**Deployment location**
Sensing and perception subsystem			Edge and fog
ITML	AV Registry	AV Registry contains metadata information of all AV sources present in MT4. These include the information on the raw streams produced by the camera and the microphone (fps rate, resolution, etc.), but also on the anonymised streams produced by the VideoAnony and AudioAnony anonymisation components.	FBK Fog WS PC (MT F2-Kubernetes)
IFAG	MEMS microphone	One MEMS microphone Nano Hub connected to a Raspberry Pi is used in MT4, specifically at Piazza Dante.	Edge microphone
MT	Camera	Two cameras with enabled streaming are used in MT4, specifically at Piazza Dante.	Edge camera
Security, privacy, and data protection subsystem			Edge, Fog, and Cloud
FORTH	EdgeSec VPN	In MT4, EdgeSec VPN creates a secure F2C VPN traffic backbone for all communications within the elements of the MARVEL platform by 100% encryption of the traffic. For MT4, the edge layer is not part of the EdgeSec VPN network.	FBK Fog WS PC (MT F2-Kubernetes) and Cloud (PSNC HPC via OpenStack)
FBK	VideoAnony	VideoAnony detects individuals that appear in the video footages from Piazza Dante and anonymises their faces. As MT4 uses one camera stream, the component is present with one instance.	FBK Fog server (MT F1)
FBK	AudioAnony+VAD	Based on the onset and offset times of speech segments detected by VAD, AudioAnony anonymises the respective segments of the	Raspberry Pi (MT4 E1)

		audio. As MT4 uses one audio stream, the component is present with one instance.	
Data management and distribution subsystem			Edge, Fog, and Cloud
ITML	Data Fusion Bus (DFB)	DFB stores inference results of the AI components that participate in MT4.	Cloud (PSNC HPC via OpenStack)
INTRA	StreamHandler	StreamHandler receives continuously anonymised AV data streams and segments them for temporary storage, for later, on-request visual inspection through SmartViz.	FBK Fog WS PC (MT F2-Kubernetes)
ATOS	DatAna Fog and Cloud	DatAna for MT4 consists of DatAna Fog and DatAna Cloud components, each continuously consuming, through MQTT, the results of AI inference components deployed at the corresponding layer (Edge/Fog/Cloud) and sending to the relevant Kafka topics at the DFB	FBK Fog WS PC (MT F2-Kubernetes) and Cloud (PSNC HPC via OpenStack)
CNR	HDD	For the predefined MT4 AI inference components, HDD optimises the allocation of their output streams across a given set of DFB Kafka topics.	Cloud (PSNC HPC via OpenStack)
Audio, visual, and multimodal AI subsystem			Fog and Cloud
GRN	CATFlow	In MT4, CATFlow classifies traffic entities (both vehicles and pedestrians) and provides their trajectories for each of the two camera streams.	FBK Fog WS PC (MT F2-Kubernetes)
AU	Audio-Visual Anomaly Detection (AVAD)	In MT4, AVAD detects anomalies in the monitored area (e.g., drug dealing, people running away, etc.).	Cloud (PSNC HPC via OpenStack)
TAU	Sound Event Detection (SED)	In MT4, SED detects audio events that occur in parking places (e.g., people shouting, type of vehicles, etc).	Cloud (PSNC HPC via OpenStack)
Optimised E2F2C processing and deployment subsystem			Cloud
FORTH	MARVdash	MARVdash provides a Kubernetes-based deployment environment for all MT4 components operating under part of the MT infrastructure managed by Kubernetes.	Cloud (PSNC HPC via OpenStack)
System outputs: User interactions and the decision-making toolkit			Cloud
ZELUS	SmartViz	SmartViz indicates the behaviour of road users and on gathering traffic statistics on road across time and the presence of anomalies in Piazza Dante.	Cloud (PSNC HPC via OpenStack)

10.1.2 Pilot E2F2C infrastructure

The MT4 infrastructure provided for the R2 integration consists of four IP cameras transmitting video, two Raspberry Pis devices transmitting audio data collected from the connected IFAG microphones, and two workstations at the fog layer, managed by FBK. The following subsections will describe each component in detail.

Edge devices

The MT4 edge layer is the same as MT2. Please refer to Section 8.1.2 of this document for more details. Table 32 lists the specifications for each device.

Table 32: MT4 Sensing devices

Specifications Type	Piazza Dante – Dogana 2, Dogana 3 and Via Pozzo	Piazza Dante – Listone 3
Video		
IP camera model	Digital cameras - Basler BIP-1600dn	Digital cameras - Basler BIP-1600c
Resolution	1600 x 1200	1600 x 1200
Frame Rate	12,5 fps	2 fps
Video Encoding	H.264	H.264
Audio		
Microphones	IFAG-MEMS (This microphone covers camera of Dogana 2 and Dogana 3)	IFAG-MEMS
Audio Encoding	ACC (LC)	ACC (LC)
Audio Sampling Rate	16kHz	16kHz
Audio Stream Bitrate	Mono 69 kbps	Mono 69 kbps
RPi	Raspberry Pi 4 Model B 8GB RAM – Micro SD 32GB	Raspberry Pi 4 Model B 8GB RAM – Micro SD 32GB

Fog workstations

FBK provides the Fog tier for the MT use cases. To comply with the constraints of the agreement between MT and FBK (FBK was nomineed as data processor) 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.

Table 30 (Section 8.1.2) lists the specifications of workstation 2. Workstation 1 is an FBK internal machine which may change and will not be accessed by users external to the research organisation.

10.1.3 Analysis of real-life data streams

In the case of MT4, video data is recorded using four IP cameras of resolution 1600x1200 pixels. Data capturing is performed using 12,5 and 2 frames per second, H.264 codec is used, and data is stored using the MP4 format.

Audio is being captured using IFAG-MEMS at a sampling frequency equal to 16kHz, whereas bit depth is 16 bits per sample. Taking into account that mono audio is recorded, the bit-rate is 69 kbps. Audio data is recorded using WAV format.

10.2 Multimodal and privacy-aware intelligence for MT4

This section describes the methods and approaches taken to collect the AV data, the datasets required for model training, the analysis of datasets the privacy assurance and anonymisation methods used in the use case.

10.2.1 Datasets for model training and privacy assurance

10.2.1.1 Datasets for model training

MT has provided data for various components. MT has provided the dataset "TrentoOutdoor – real recording" and "TrentoOutdoor – staged recording" (as defined in D2.1) processing the streams of cameras and microphones from Piazza Santa Maria Maggiore.

MT, in collaboration with FBK, collected, anonymised and annotated more than 150 audio and video files, in total 25 GB used for the training of:

- SED – MT annotated more than 100 raw audio samples which correspond to a total of 2 hours. The size of the samples is 0.5 GB. The samples contain both anomalous and non-anomalous situations. The files were annotated using the ontology needed to train SED;
- AVAD – MT annotated 25 GB of video streams which contain anomaly and not anomaly situations.

Thanks to the staged recording done during M16, M28, and M29, MT and FBK have provided 30 audio-videos, each manually annotated and used for training of SED and AVAD.

10.2.1.2 Analysis of datasets

The analysis of the streams and datasets is performed by the AI model providers. Pending feedback, more data for training will be added.

10.2.1.3 Privacy assurance and anonymisation

The collected video data are anonymised at FBK premises using the VideoAnony tool. Instead, the collected audio data are anonymised by the composite component AudioAnony+VAD on edge devices (Raspberry Pis). In this way, MT and FBK comply with the privacy constraints defined in the appointment of FBK as data processor, and they can manage and share data with consortium partners.

For the staged recordings, anonymisation is not necessary as all involved subjects signed the informed consent.

10.3 Demonstration

10.3.1 The Decision-making Toolkit

In the MT4 use case, the dashboard is designed to be used by both policy-makers/mobility managers and by local police officers, providing them with a single platform (Figure 5) to combine information and explore the available data and analysis.

The detections are visualised over time to enable monitoring and clustering using the Temporal Representation widget. This widget visualises the data from all the included components in this use case. The Vehicle Trajectories widget uses the CATflow component to draw the paths of passing vehicles in the camera feed image. The paths are grouped and colour-coded based on different vehicle types. Users can change the time period and filter by vehicle type to conduct further investigations.

Detailed information about the detected incoming events is presented in a tabular form in the Details widget. This widget also supports standalone text filtering to facilitate quick searches through the available information.

Within the audio and video player functionality in SmartViz, users can select an event and play the corresponding audio or video snippet. The StreamHandler component retrieves the stream for the relevant time period of the detection, segments it, and plays the snippet in the Audio or Video player widget within SmartViz.

Users have the ability to validate the inference results by marking them as accurate or not. The use case also includes a police intervention functionality. After reviewing the events, users can mark an event as important and in need of police intervention, enabling appropriate action.

In addition to rearranging and resizing widgets, the dashboard view provides the functionality to download the visualised data in JSON format. Users also have the option to save the entire dashboard as a PDF file, allowing for offline access and convenient sharing based on individual preferences.

Finally, the Weather Information widget in the use case provides users with a representation of weather-related data. It allows them to view and explore weather information for a selected time period, which can help uncover correlations between detected events and anomalies and weather conditions.

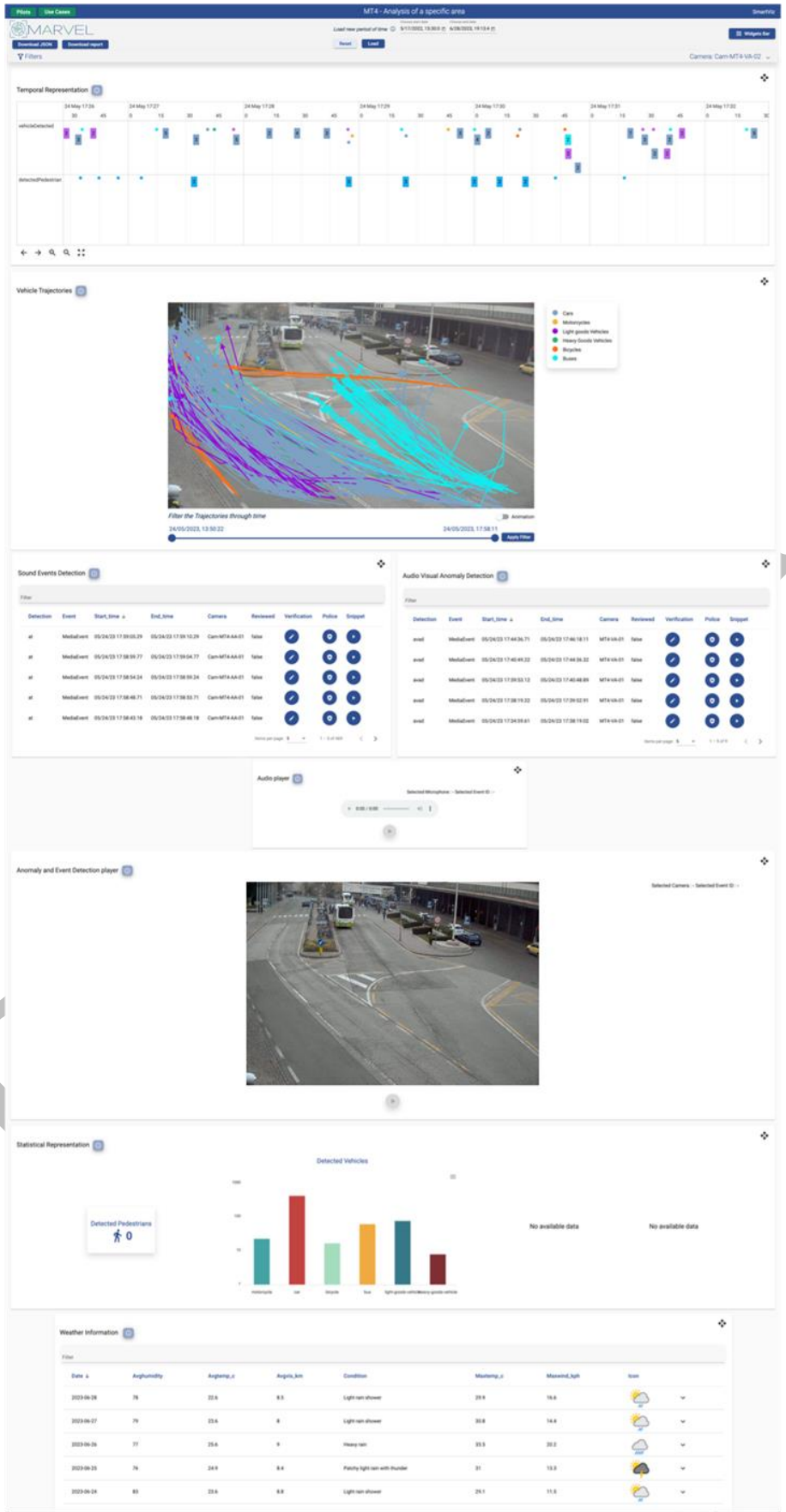

Figure 5. MT4 in the DMT

D4.6 reports the complete functionalities, widgets, and users requirements regarding DMT in MT4.

10.4 Work carried out

For a detailed description of the work done, please refer to Section 8.4 of this document.

11 UNS1: Drone Experiment

UNS1 use case is briefly described in Section 2.3.5 of this document.

11.1 Updates compared to R1

The UNS1 use case received updates based on constructive feedback provided to the ZELUS team. These updates were aimed at improving the user interface and enhancing the presentation of data in a more user-friendly manner:

- the way crowd counting maps are handled from the technical view point;
- download option of a snapshot of the system state, logs, etc. as evidence of what has occurred in the system, which can be shown to the police or local authorities was introduced.

The UNS1 infrastructure was updated including the addition of EdgeSec TEE as the secondary element responsible for enhancing security and privacy aspects. In addition, one RPI device was changed, due the malfunctioned, but it was replaced with a new RPIv4 device.

All the infrastructure changes were extensively tested.

12 UNS2: Localising Audio Events in Crowds

This section describes the steps taken to integrate the components such that the UNS2: Localising Audio Events in Crowds use case could be set up and tested. The following sections describe the framework configuration and data streams, the multimodal and privacy-aware intelligence applied, the demonstration of the implementation realised, and the work carried out.

12.1 10.1 Framework configuration and data streams for UNS2

12.1.1 Components

The implemented components in UNS2 are summarised in Table 33.

Table 33: MARVEL components in the UNS2

UNS2: Localising Audio Events in Crowds			
Component owner	**Subsystem /Component**	**Comments on how the component is used in UNS2 for R2**	**Deployment location**
Sensing and perception subsystem			Edge and Fog
ITML	AV Registry	AV Registry contains metadata information of all AV sources present in UNS2. These include the information on the raw streams produced by the microphone board (sample rate, bit-rate, etc.), but also on the anonymised streams produced by the AudioAnony anonymisation components.	UNS Fog server (UNS F1)
IFAG	Dual-PCB 8-microphone Audiohub Nano board	In UNS2, one 8-channel microphone board is used to provide an audio recording of simulated crowds.	Edge microphone
Security, privacy, and data protection subsystem			Edge, Fog, and Cloud
AUD	VAD	VAD detects speech segments in an audio stream and outputs the respective onset and offset times.	UNS Laptop (UNS E3)
FBK	AudioAnony	Based on the onset and offset times of speech segments detected by VAD, AudioAnony anonymises the respective segments of the audio. Although UNS2 records 8ch audio stream for sound event localisation and detection, only stream from one channel is enough for human check as the microphones are placed near the board. Thus, one channel is anonymised and streamed to the Fog.	UNS Laptop (UNS E3)
Data management and distribution subsystem			Edge, Fog, and Cloud
ITML	Data Fusion Bus (DFB)	DFB stores inference results of the AI components that participate in UNS2.	Cloud (PSNC HPC via OpenStack)
INTRA	StreamHandler	StreamHandler receives continuously anonymised audio data streams and segments them for temporary storage, for later, on-request visual inspection through SmartViz.	UNS Fog server (UNS F1)

ATOS	DatAna Edge, Fog and Cloud	DatAna for UNS2 consists of DatAna Fog and Cloud agents, each continuously consuming, through MQTT, the results of AI inference components deployed at the corresponding layer (Edge/Fog/Cloud) and sending to the relevant Kafka topics at the DFB. DatAna Edge consists of MQTT only, while DatAna Fog and Cloud implement also the NiFi functionality. The inference results gathered at the edge and fog layers are therefore consumed and processed by the NiFi located at the fog from MQTT (edge and fog), while the results produced at the cloud are consumed by the NiFi at the cloud from the MQTT deployed at the same layer.	UNS Fog server (UNS F1) and Cloud (PSNC HPC via OpenStack)
Audio, visual, and multimodal AI subsystem			Fog
TAU	Sound Event Localisation and Detection (SELD)	In UNS2, SELD component recognises targeted sound events and provides their spatial trajectories.	UNS Laptop (UNS E3)
Optimised E2F2C processing and deployment subsystem			Fog and Cloud
FORTH	MARVdash	MARVdash provides a Kubernetes-based deployment environment of all the UNS1 components.	Cloud (PSNC HPC via OpenStack)
System outputs: User interactions and the decision-making toolkit			Cloud
ZELUS	SmartViz	SmartViz visualises the output of the SELD component, i.e., presents trajectories of detected sound events on the city map, where positions of the microphone boards used are predefined.	Cloud (PSNC HPC via OpenStack)

12.1.2 Pilot E2F2C infrastructure

The UNS2 use case infrastructure consists of the Edge, Fog, and Cloud tiers. An IFAG Audiohub – Nano 8-channel microphone board is used as a data capturing device at the edge. It is connected to the laptop at the edge through Wi-Fi network. The microphone board captures audio data using eight spatially distributed sensors of the microphone array and transmits it to the laptop where SELD and AudioAnony components are deployed. The user can set the sampling rate by selecting from the predefined set of values as well as bit-rate and number of channels for data capturing (up to 8 channels).

Default parameter values of the microphone board are shown in Table 34 whereas specification of the laptop is presented Table 35.

Table 34: UNS2 microphone specification at the Edge

Specifications Type	Specifications
Microphone array mode	Dual-PCB 8-microphones Audiohub – Nano board
Channels	8
Sampling rate	48kHz
Rate	24-bit audio data

Table 35: UNS2 laptop at the Edge

HW subsystem	Specifications
Laptop model	HP ProBook 440 G9
CPU	Intel Core i7 - 1255U - 4.7GHz 10 cores, 12 threads, 12MB cache
GPU	Integrated
Hard Drive	1 TB
RAM	16 GB

The description of the Fog server is presented in Table 36. It serves for running AVRegistry, StreamHandler, and DataAnaFog. All infrastructure components are locally connected with multiple redundant 1G Ethernet links. UNS infrastructure is currently hosted "behind" an HTTP(S) proxy for Internet connectivity.

The Cloud tier serves for the model training and visualisation which is enabled using the SmartViz component. SmartViz provides users with a city map of deployed microphone boards with a visualised direction of the detected and localised audio events. Other components deployed at the Cloud are DFB and DatAna Cloud.

Table 36: UNS2 devices at the Fog

HW subsystem	Specifications
Server model	SUPERMICRO SYS-7049A Server SuperWorkstation
CPU	2 x Intel Xeon Silver 4110 - 2.1 GHz 8 cores, 12 threads, 11MB cache
GPU	Nvidia TitanXP
Hard Drive	3 x 300GB SSD and 1 x 1TB SSD
RAM	128GB DDR4

12.1.3 Analysis of data streams

The microphone used for recording is IFAG Audiohub – Nano 8 Mic. This audio array provides recordings from 8 channels at 24 kHz and resolution of 24 bits per sample. The data from all 8 channels is streamed to the laptop where the audio event localisation is performed. The anonymised stream from one of the microphones is sent to the server such that a human user is able to confirm the events that occurred through the audio.

12.2 Multimodal and privacy-aware intelligence for UNS2

This section describes the methods and approaches taken to collect the AV data, the datasets required for model training, the analysis of datasets, and the privacy assurance and the anonymisation methods used in the use case.

12.2.1 Datasets for model training and privacy assurance

12.2.1.1 Datasets for model training

UNS has collected audio data using IFAG Audiohub – Nano 8 Mic. Three types of target events were extracted from the FSD50k dataset[9]: boom, gunshot, gunfire and shatter. Since the provided audio files are a bit longer than the target event itself, the files were first semi-automatically processed to extract only audio parts of interest.

The audio was reproduced using eight JBL VP7212MDP[10] speakers, which were positioned circularly around the microphone in equidistant positions. One of the speakers was used to reproduce target events, while the others were used to reproduce background noise. The 10-second background noise samples were randomly chosen from the FSD50k dataset and belong to "chatter" sound type. The target events were reproduced exactly in the middle of 10-second segments.

There were two recording setups. In the first, the speakers were positioned at a 5m distance from the recording microphone array, while in the second, the distance was 10m. For both recording setups, all the extracted target events were reproduced but the speaker reproducing target event was randomly chosen for each setup. The recording process was split into 6 sessions, each session lasting approximately 75 minutes. The default microphone sampling rate of 48k was used.

Besides the IFAG microphone array, which is used for data collection in UNS2, an additional calibrated microphone was used within the staged recording. The additional microphone was positioned at the same place as the data capturing device in order to estimate the overall sound pressure level at the recording position.

The recording was performed at the property of Studio Berar, who provided us with all the recording equipment, in Novi Sad[11]. All recordings were taken between 10 am and 8 pm.

12.2.1.2 Analysis of datasets

The annotated UNS2 dataset is used for training SELD. In the dataset postprocessing, the recorded mixtures (10-second segments containing target event and background noise) were automatically extracted and manually checked. The number of extracted mixtures is shown in Table 37. These numbers represent the overall number of mixtures for both recording setups. The size of the dataset is approximately 34.3 GB. The train/validation/test split was performed based on the information from FSD50k. The target event added to one set (train, validation or test) in FSD50k is associated with the same set in the recorded database. For each mixture, the label is automatically created containing all information about target events, as well as background noise tracks.

Table 37: Database content

	Train	Valid	Test
Boom	280	**54**	**74**
Gunshot	588	**108**	**268**

[9] Fonseca, E., Favory, X., Pons, J., Font, F. and Serra, X., 2021. Fsd50k: an open dataset of human-labeled sound events. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 30, pp.829-852

[10] https://jblpro.com/en/products/vp7212mdp

[11] https://studioberar.com/en/contact/

Shatter	700	128	192

12.2.1.3 Privacy assurance and anonymisation

Staged recordings were conducted by UNS using samples from the open database FSD50K, which were prepared for 8-channel audio processing, simulating anomalous events in the crowds. While the data can be processed without anonymisation, as the recording did not involve humans, UNS2 incorporates an audio anonymisation component to meet the needs of the envisioned end-users that would monitor a real crowd. In the case of audio processing, two components were used - VAD to detect voice segments as well as the AudioAnony component to anonymise audio segments when a voice is present. All anonymisation is carried out at the edge.

12.3 Demonstration

12.3.1 The Decision-making Toolkit

The objective of the UNS2 use case is to investigate the feasibility of localising and detecting audio events within crowds using audio streams.

The final dashboard for UNS2, as depicted in Figure 6, includes several key features. One of the prominent features is the Sound Localisation map, which visualises the outputs of the SELD component. Users can interact with this widget to observe consecutive detected events within a selected time period. Each event is represented by an arrow indicating its direction.

Furthermore, the detected events, along with the output of the VAD component, are displayed in a details widget. Users can request the corresponding audio snippet of an event and play it in the audio player.

To ensure the accuracy of the inference results, users have the ability to validate them as either accurate or inaccurate. The dashboard provides flexibility with the option to rearrange and resize widgets according to the users' preferences.

Additionally, the dashboard view offers functionality for users to download the visualised data in JSON format. Furthermore, users can save the entire dashboard as a PDF file, enabling offline access and facilitating convenient sharing based on individual preferences. These features enhance the usability and accessibility of the dashboard.

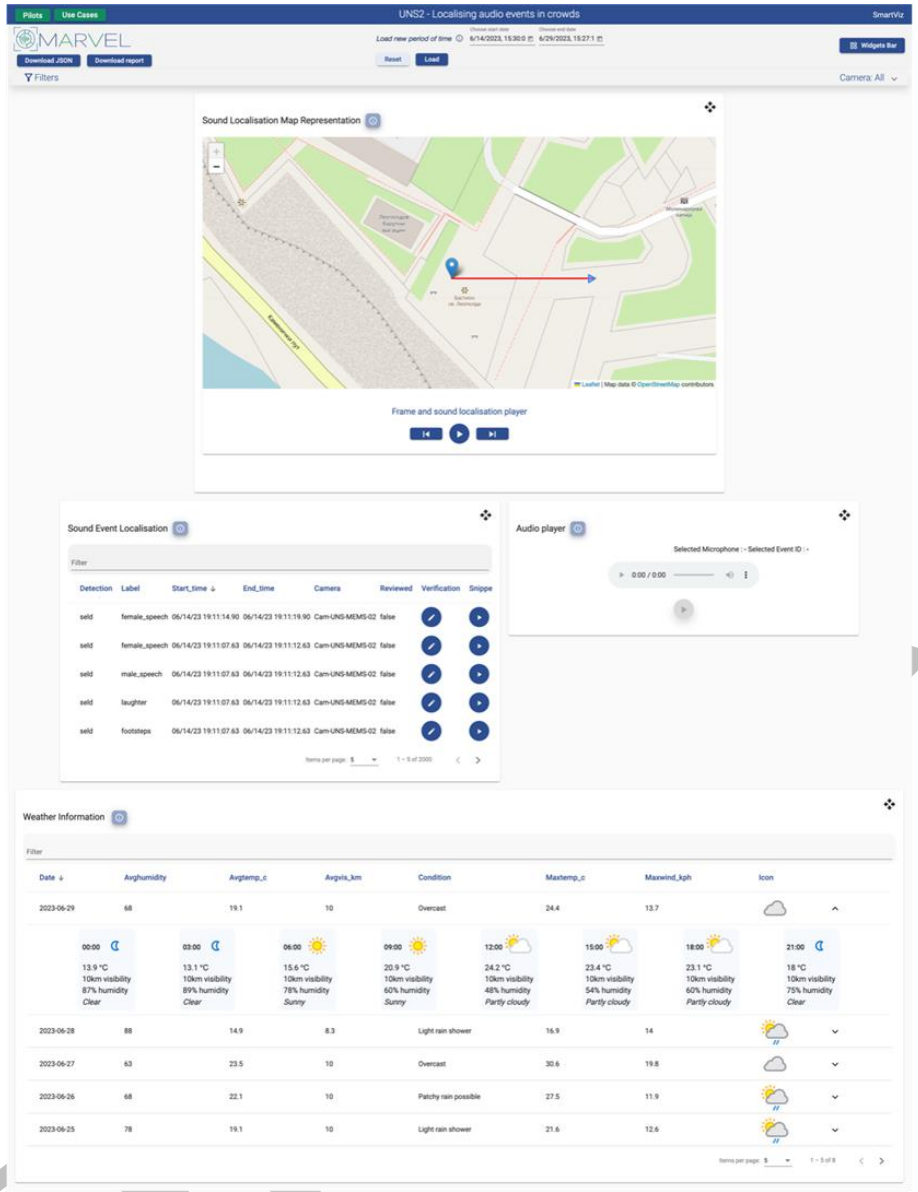

Figure 6. UNS2 in the DMT

D4.6 presents the complete functionalities, widgets, and users requirements regarding DMT for UNS2.

12.4 Work carried out

Development of the UNS2 use case was carried out in two phases. Firstly, the Audio-Visual Emotion Recognition use case was defined. UNS made an initial architecture definition and the first phase of database recording was performed. Four participants were involved in the acting of five emotions, which were recorded and about 80 minutes of audio-visual data were collected. In addition, the privacy-preserving audio-visual emotion recognition component was developed, as it is described in D3.5.

Despite the efforts made the use case was abandoned and redefined, in M26 as it was described in Section 2.2.3 due to legal concerns. Several meetings between UNS and PN with a focus on legal aspects of the use case preceded the decision as well as discussion with the whole consortium.

In the early phase of the use case redefinition, several meetings were held between UNS and TAU with a focus on the experimental design as their component, SELD, was crucial for further development within the project. After that, UNS organised two focused meetings with IFAG in order to perform the implementation of their AudioHub Nano 8 channel microphone arrays into the experimental setup. UNS has also organised bi-weekly pilot-focused meetings after the decision to redefine UNS2 use case. The initial efforts, deciding to redefine the use case, included detailed planning of staged recording procedure and architecture definition, whereas reports of the E2E tests and visualisation refinements were performed in the later phase.

The staged recording process was performed in cooperation with a professional music studio Berar from Novi Sad since professional audio equipment was required for quality recording of sound event localisation and detection dataset. Recorded data were prepared using samples from the FSD50K database, as it was described in Section 12.2.1.1.

13 Conclusions

This report presented the final version of the MARVEL Demonstrators' execution. The following results are reported in this deliverable: (a) the revision of the definition of the experimental protocol taking into account the outcomes of the implementation of the MVP at M12, the first version of the system at M18, and any new insights derived from the data collections; (b) the shared definition and consequently the selection and implementation of the aspects that characterise all the selected use cases of the final release; and (c) the finalisation of coordinated activities with the aim of improving the use cases realised for M18 and implementing the selected use cases for M30 in terms of configuration of the framework and data streams, selecting multimodal and privacy-aware intelligence applied and analysing the results and demonstrating their application. All these achievements confirmed the versatility of the MARVEL framework and its components.

The application and integration of the tools provided by the MARVEL framework had to deal with real-life situations, which were often not straightforward, especially due to the restrictions related to privacy and the deployment of the technological components used for data collection.

The experience gained serves as a starting point for refining all the considered use cases. This refinement encompasses enhancing functionality based on feedback received from end users. Additionally, the insights gained during this process are instrumental in benchmarking performance and identifying potential opportunities for commercialisation activities.

In conclusion, this document can be used as a basis and reference for the upcoming planned activities, deliverables and milestones, especially for those related to the final evaluation of the execution of the MARVEL Demonstrator (D6.4 due M36).

14 Appendix

Emotion recognition artificial intelligence

Focus on usefulness and ethical issues

AI is a transformative field of research that focuses on endowing various systems with the ability to perceive and interpret human emotions, among other things. AI that is capable of recognising emotions, has enormous applications in a variety of fields, from e-commerce and marketing to education and healthcare. However, as the potential of this technology unfolds, it is essential to address ethical challenges related to data protection and personal privacy. Therefore, this analysis details the general utility of artificial intelligence capable of recognising human emotions. It also outlines the ethical challenges associated with the development and use of this technology from a personal privacy and data protection perspectives.

Usefulness of emotion recognition AI

AI developed for the purpose of recognising emotions improves the interaction between humans and machines. This allows machines to understand human behaviour and respond in ways that personalise the user experience, make recommendations, and increase user satisfaction. Improved user experience is significantly important in customer-facing industries such as retail, e-commerce, and hospitality. In these industries, timely and appropriate responsiveness positively impacts personalised experiences by addressing customer needs, providing timely support, and improving customer satisfaction and loyalty. In addition, emotion recognition AI can support market research and advertising by evaluating consumer responses. Analysing emotional responses to advertising, product designs, or marketing campaigns can provide valuable insights into consumer preferences and help companies develop more effective strategies and targeted marketing campaigns.

AI systems capable of recognising emotions can be an important factor in supporting health. They can help monitor patients' well-being by analysing emotional signs and detecting signs of distress or discomfort. Analysis of facial expressions, voice, or even body language can detect and evaluate mental states. As such, this technology can help healthcare professionals provide better care to patients and improve overall well-being. In addition, its application could be incorporated into diagnostic techniques and methodology.

The usefulness of AI for recognising emotions is also evident in human resource management. Signs of personal emotions can reveal factors that are useful for selecting job candidates and making hiring decisions. It can also be used when analysing employee well-being and satisfaction. Concerning education, emotion recognition can help personalise learning experiences. By monitoring students' emotional state, the system can adjust teaching methods, provide feedback, and offer support tailored to individual needs. This can improve student engagement, motivation, and overall learning outcomes.

Finally, AI capable of recognising emotions can be used in security and safety systems. It can help detect suspicious behaviour or potential threats by analysing facial expressions or behavioural patterns. This technology can improve surveillance systems, enhance public safety, and help with early threat detection.

Ethical challenges (focus on data protection)

AI developed for the purpose of emotion recognition poses several privacy and data protection challenges. The most noticeable challenge can be found in processing of special categories of data, which essentially deserve a special kind of protection. Algorithms may reveal certain personal identities, such as those related to health status (assuming that certain mental states might be related to health status). For example, some severe psychiatric and neurological disorders such as psychosis could be detected by observing emotions. Also, analysing historical data on a person's emotional state may reveal other health conditions such as depression. Not to be neglected, even political opinions or philosophical beliefs can be determined by interpreting emotions and their signs. The use of such sensitive personal data is especially dangerous when processed by unauthorised entities. AI capable of recognising emotions, however, essentially does not just identify a person; its primary goal is to interpret emotions, analyse their accuracy, and decide on further application elements.

Technologies that recognise human emotions belong to the group of disruptive technologies. These technologies convert emotional signals into information that is used to make specific decisions. Careful consideration must be given to whether the processed data is necessary to achieve the intended goals. It is also necessary to consider whether the same goal can be achieved with a less intrusive alternative. This type of evaluation is known as the necessity and proportionality test and must be performed on a case-by-case basis (if the regulatory requirements are strictly obeyed). However, it is more than obvious that an assessment for each individual case where data is processed in the context of use of these technologies is quite difficult (or even impossible) to perform. The assessment of proportionality and necessity should take into account many factors, such as the nature of the data collected, the type of the inferences, the purpose of the processing, the duration of the data retention and the like.

Facial expressions and other signs of emotion may vary from person to person if they express different emotional states (someone may be laughing when happy while someone else is crying). Sometimes the observation of a limited number of signs (e.g., from facial expressions) can be used to infer a particular state, while the combination with another group of signs (and information), e.g., from body language, can give a different impression about an emotional state. Therefore, observing only one group of data can lead to false impressions. Also, we should not neglect that humour, sarcasm, or irony are influenced by the sociocultural context. In addition, technical aspects of information collection may influence the quality of processed information. Therefore, accuracy is brought into play as one of the principles of personal data protection in the context of AI development. Accuracy may be limited, and consequently, inputs that affect a person's life could lead to undesirable scenarios.

Problems with data accuracy can also lead to discriminatory scenarios. As mentioned, sociocultural patterns can affect the context of interpretation of emotional signals, and if the patterns are processed inappropriately, bias based on ethnicity or skin colour can occur. In addition, if the AI system does not take into account the context of a person's transient situation (e.g., a person's behaviour when ill or injured), misclassification of certain groups is likely. This leads to potential discrimination, i.e., a confrontation with the principle of fairness. Therefore, selecting the right data set that has the characteristics of a representative sample (or samples) is essential to avoid discrimination and maintain fairness.

The following challenge relates to the preservation of transparency and ensuring lawfulness. Once AI is developed to recognise emotions, it must be used in the context of its ultimate purpose (which may be commercial, educational, national security, or otherwise). Capturing emotional signs, including facial expressions and/or body language, is not technically difficult

(images and video can be captured anywhere due to the ubiquity and small size of cameras), but it is quite a challenge to ensure that the capture is done lawfully. In many situations, people should give their consent to be recorded. However, in real life, it is impossible to obtain the consent of everyone involved in mass surveillance or when data is processed on a large scale. Therefore, alternative legal grounds for ensuring lawful data processing should be considered. Another issue related to the lawfulness of data processing is how to inform individuals that they are being recorded and the ultimate impact of recording on the use of AI systems.

Individuals must be accurately informed, and they must be able to access and control the use of AI. Otherwise, they will be deprived of the freedom to choose which aspects of their lives can be used to influence the context of emotion management. In addition, advanced AI algorithms are quite complex, and therefore providing information about this complexity in a concise and easy-to-understand manner to an average member of many populations targeted for processing is a challenging task.

Notwithstanding the numerous benefits of AI for emotion recognition, the fact is that the data subject may not be aware of the use of these profound systems and potential profiling and targeting processes. Namely, these systems can be used to create profiles of people in a number of situations. Inference based on profiling is based on the association of an individual with a certain group of people. Thus, emotion-based experience (processed by algorithm) is supposed to be the ultimate benchmark for grouping people into predetermined groups of individuals. Obviously, this decision-making logic could result in false-negative and false positives that further lead to potential of discrimination and other adverse consequences.

Instead of conclusion

AI developed to recognise emotions is a promising tool that could revolutionise human-machine interaction. It can facilitate personal experiences in various fields and improve the economy, business, health care, education, and public safety. Nevertheless, ethical challenges related to data protection and privacy must be addressed appropriately and proactively. It is important to ensure that the use of emotion recognition systems is accompanied by responsible practices that respect the rights of individuals and prioritise the well-being of users. This can only be achieved if fundamental ethical principles are respected.