# 2021 BenchCouncil International Symposium on Benchmarking, Measuring and Optimizing (Bench'21)

Nov 14-16, 2021

## Tutorial-4:
## DataBench Toolbox for Pipeline-based selection of Big Data and AI Benchmarks

MARVEL

Multimodal extreme scale analytics for smart cities environments

*Example of application of AI and Big Data Pipeline Framework and benchmark identification in the MARVEL project*

Dragana Bajovic

# Identity Card

**Project Consortium**
17 partners

**Project Name**
Multimodal Extreme Scale Data Analytics for Smart Cities Environments

**Total Budget**
€5 998 086.25

**Project Type**
RIA

**Start Date**
1 January 2021

**Duration**
36 months

https://www.marvel-project.eu marvel-info@marvel-project.eu @MARVELprojecteu MARVEL project

# Motivation

Cities have become actual **“data engines”**

*huge variety of IoT urban sensors and devices recording multiple everyday activities producing large scale heterogeneous datasets*

Need for accurate **predictions** and better **analytics**

**Challenges**

1. Valuable **knowledge extraction**
2. **Commercial value** from data

Need to **shift** traditional **methodologies**, **techniques** and **tools** of information extraction into **new dimensions.**

**How?**

By cracking the problem of **extreme scale data analytics**

# Our Vision

*MARVEL* aims to foster the vision of EU Data Economy by:

1. **addressing** and solving **challenges** in the **Big Data Value chain**
2. prioritising strengthening of **open science** and **open data** through enriching and sharing a **Data Corpus** to drive R&I
3. heavily **investing** in **R&I** to derive **new knowledge** and advance existing one, ensuring a **sustainable growth** for the technological advancements
4. engaging **citizens** to promote breakthrough **innovation.**

# MARVEL Framework - Pillars

**1** Real **heterogeneous distributed** Big Data in **smart cities** environments

**2** **Edge-to-fog-to-cloud** (E2F2C) distributed ubiquitous computing **architecture**.

**3** **AI-based intelligence** for **multi-modal perception** and **situational awareness.**

**4** **Quantitative assessment** of E2F2C and **Multi-modal AI tools and methods** via societal, academic and industry validated **benchmarks**.

# MARVEL's conceptual architecture

# MARVEL's Data Value chains

**Data Acquisition/Collection**
(including Data Ingestion processing, Streaming, Data Extraction, Ingestion Storage, Different data types)

**Data Storage/Preparation**
(including Storage Retrieval/Access/Queries Data Protection, Curation, Integration, Publication)

**Analytics/AI/ Machine Learning**
(including data analytics, ML model training, model verification, operation

**Action/Interaction, Visualisation/Access**
(including data presentation environment/boundary/user action and interaction)

# The MARVEL Smart City test cases

**MT**

Monitoring of crowded areas
Detecting criminal and anti-social behaviours
Detection of possible dangerous situation
Monitoring of parking places
Analysis of a specific area

Trento

**UNS**

Drone experiment
Audio-visual emotion detection

Novi Sad

**GRN**

Safer roads
Road user behaviour
Traffic conditions and anomalous events
Junction traffic trajectory collection

Malta

# Novi Sad Pilot: Experimental data collection

AVDrone

- Raspberry Pi Camera Module 8MP v2.1
- Infineon MEMS microphone
- Raspberry Pi board
- Intel NUC miniPC
- Wi-Fi and LTE-M connectivity
- DJI Matrice 600 Pro drone

UNS Audio - visual emotion recognition

- Android app for recording

UNS will perform staged recordings for both use cases and all participants will sign a written informed consent.

# Novi Sad Pilot: Data collection and processing examples

(a)

(b)

(c)

*Hassouneh, A., Mutawa, A. M., & Murugappan, M. (2020). Development of a real-time emotion recognition system using facial expressions and EEG based on machine learning and deep neural network methods. *Informatics in Medicine Unlocked*, *20*, 100372.

(a) Drone experiment recording place (UNS campus)
(b) Exit festival (potential end user)
(c) Emotion recognition example

# Novi Sad (UNS) use case: Drone experiment

# Trento Pilot: introduction

- With its **120.000 inhabitants**, Trento is the third-largest city located in the heart of the Alps and the capital of the Province of Trento
- The Municipality of Trento employs about **1.500** collaborators
- As one of the Smart Cities in Italy [1], Trento is currently focused on finding a **smart, privacy-preserving way** to analyse the vast amount of **data** collected from the different devices and sensors in the city
- **Objective**: support an effective and pragmatic decision making process based on specific situation or context

[1]https://www.forumpa.it/citta-territori/icity-rank-2020-firenze-bologna-e-milano-sono-le-citta-piu-digitali-ditalia/

# Trento Pilot: issues and solution

## Current issues in TRENTO

- The Municipality of Trento (MT) wants to understand what is happening in various outdoors areas of the city (parks, squares, underpasses and crossroads)
- Municipality of Trento owns hundreds of **Cameras** (more than 600), but the staff cannot handle the amount of data, its velocity, and its heterogeneity
- Currently, the data are only used a posteriori

## MARVEL solution

- Real-Time analysis + alert
- AI algorithms → prediction of danger
- Smart interface for the Local Police

# Trento Pilot: use cases and data collected

**Use Cases:**

I - Monitoring of crowded areas
II - Detecting criminal and antisocial behaviours
III - Monitoring of parking places
IV - Analysis of a specific area

**Situation that will be detected:**

- gatherings
- robberies
- aggressions/fights
- drug dealing
- car accidents
- collection of data (number of persons, cars, trajectories, events) in order to support long term decision making by public authorities

**Notification system** will be used for:

- alerting the operational centre of local police station via an alarm
- creating a custom view on a smart interface to highlight the relevant cameras

**Potential end users:**

- Police
- Armed forces
- Municipalities
- Private companies with public participation (managing parking areas, bike sharing, car sharing, etc.)

# Trento (MT) use case: Monitoring of Crowded Areas

# Malta Pilot: Introduction

## Motivations

- The Malta National Transport Strategy[1] recognises:
  - Planning needs to be carried out over an analytical and scientific basis,
  - The need to encourage walking and cycling
  - The importance of education campaigns.
- Local studies[2] strongly indicate that the lack of safe roads is the main barrier to cycling

[1]National Transport Strategy 2050 Malta, https://www.transport.gov.mt/strategies/strategies-policies-actions/national-transport-strategy-and-transport-master-plan-1343

[2]Maas and Attard 2021, Policies to promote cycling in Southern European island cities: challenges and solutions from three 'starter' cycling cities, Transportation Research Procedia, 2021

# Malta Pilot: Introduction

## Overall goals

- Encourage uptake of active modes for commuting (e.g. Cycling and micro-mobility)
- Collect real-time data of traffic state for short-term decision making
- Collect data for long-term transport analytics  (mainly for upgrading the infrastructure)
- Support campaigns aimed at educating road users

## Problems to be solved

- Detecting cyclists (from V or AV data)
- Detecting anomalies in AV data (e.g queues, obstacles, accidents)
- Collection of trajectories of road users as they move across junctions or road sections.

## End-users

- Vulnerable road users (Cyclists and pedestrians)
- Traffic Engineers and Traffic Control Personnel
- Transport researchers and Policy makers

# Malta Pilot: Data collection and processing examples

(a)

(b)

(c)

Example locations for data collection and task execution in Malta:

(a) inter-city link,
(b) urban junction,
(c) sub-urban junction

# Malta (GRN) use-case: Safer Roads

# Benchmarking continuum

*<u>Project's rationale</u>: for each of the project's components, functionalities, and end-to-end applications, perform a systematic benchmarking.*

**Setup**

introduces societal, academic and industry validated benchmarks, and define strategy (business, technical and user experience) to continuously update them

**Testing**

**defines the appropriate test cases to drive the technical evaluation of the framework, according to the benchmarks introduced in the Setup**

**Evaluation**

provides evaluation and KPIs monitored for each experiment both in operational (cost, service levels etc.) and technical terms (performance of solution)

- *Business (model) Benchmarking*, which will validate the *MARVEL* process, sustainability and financial perspectives.
  - European Data Market Monitoring Tool
  - BDVe Landscape, DataBench Toolbox
- *Technical Benchmarking*, which will determine the quality and performance of *MARVEL* technologies
  - DataBench Toolbox, Benchmarking communities
- *User Experience Benchmarking*, which utilizes user-centric indicators (e.g. usability and innovation) to validate the quality of end user's experience.

# MARVEL's Data Value chains

**Data Acquisition/Collection**
(including Data Ingestion processing, Streaming, Data Extraction, Ingestion Storage, Different data types)

**Data Storage/Preparation**
(including Storage Retrieval/Access/Queries Data Protection, Curation, Integration, Publication)

**Analytics/AI/ Machine Learning**
(including data analytics, ML model training, model verification, operation

**Action/Interaction, Visualisation/Access**
(including data presentation environment/boundary/user action and interaction)

**GDPRBench**

**HiBench, SparkBench RIoTBench, TPCxIoT**

**TPC-DS v2**

***Cloud***: HiBench, YCSB
***Fog***: TPCx-IOT, RIoTBench, DeFog
***Edge***: Edge AIBench, TPCxIoT, DeFog
***E2F2C***: LEAF, DeFog
Benchmark of Regex libr.

# Thank you!

This project has received funding from the European Union's Horizon 2020 Research and Innovation program under grant agreement No 957337.

https://www.marvel-project.eu marvel-info@marvel-project.eu @MARVELprojecteu MARVEL project