DATAWEEK

JOIN.LEARN.SHARE.GET VALUE

Big Data and AI Pipelines in Big Data PPP projects from a Technology analysis and benchmarking perspective

Project MARVEL: Multimodal extreme scale analytics for smart cities environments

May 26th, 11:30-13:30

#hashtag

EUHUBS4DATA

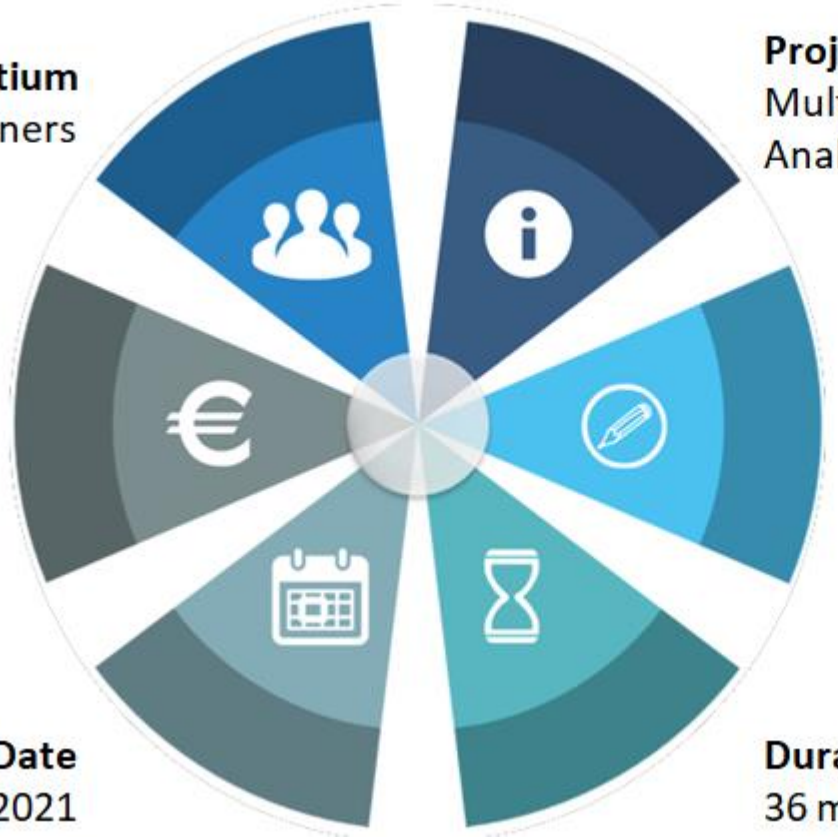
BDV BIG DATA VALUE ASSOCIATION

DECSIS ubiwhere valuedate.io

Identity Card

Project Consortium
17 partners

Project Name
Multimodal Extreme Scale Data Analytics for Smart Cities Environments



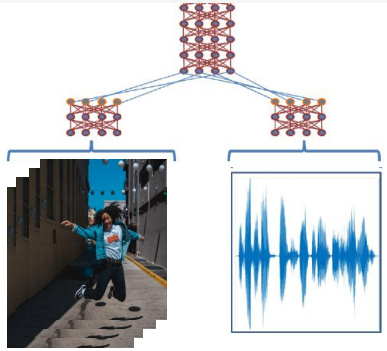
Total Budget
€5 998 086.25

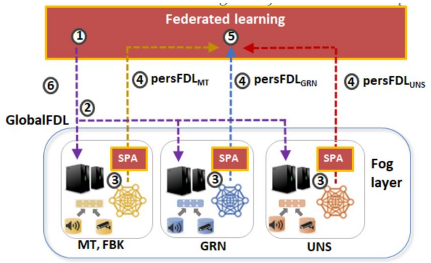
Project Type
RIA

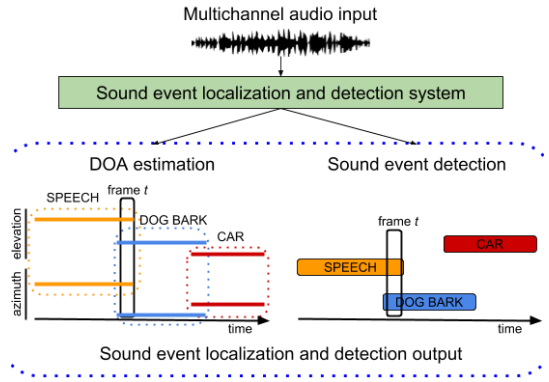
Start Date
1 January 2021

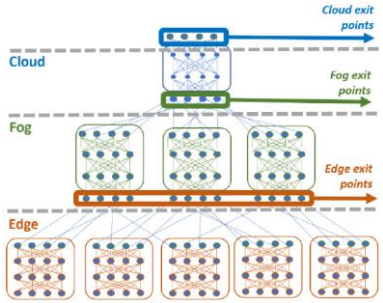
Duration
36 months

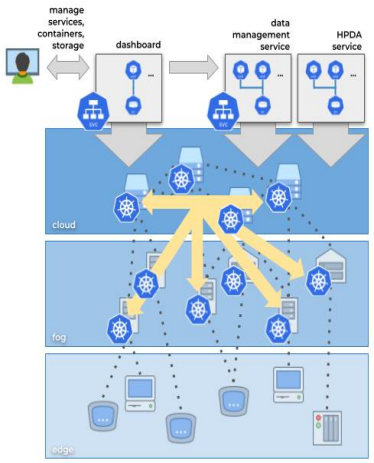
MARVEL Framework - Pillars











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**.

The MARVEL Smart City test cases

MT

Monitoring of crowds
Detecting criminal/unsocial behaviours
Detection of possible dangerous situation
Monitoring of parking places/subway
Analysis of a specific area

Trento

UNS

Crowd classification and anomaly detection for monitoring open space public events
Audio-visual emotion detection

Novi Sad

GRN

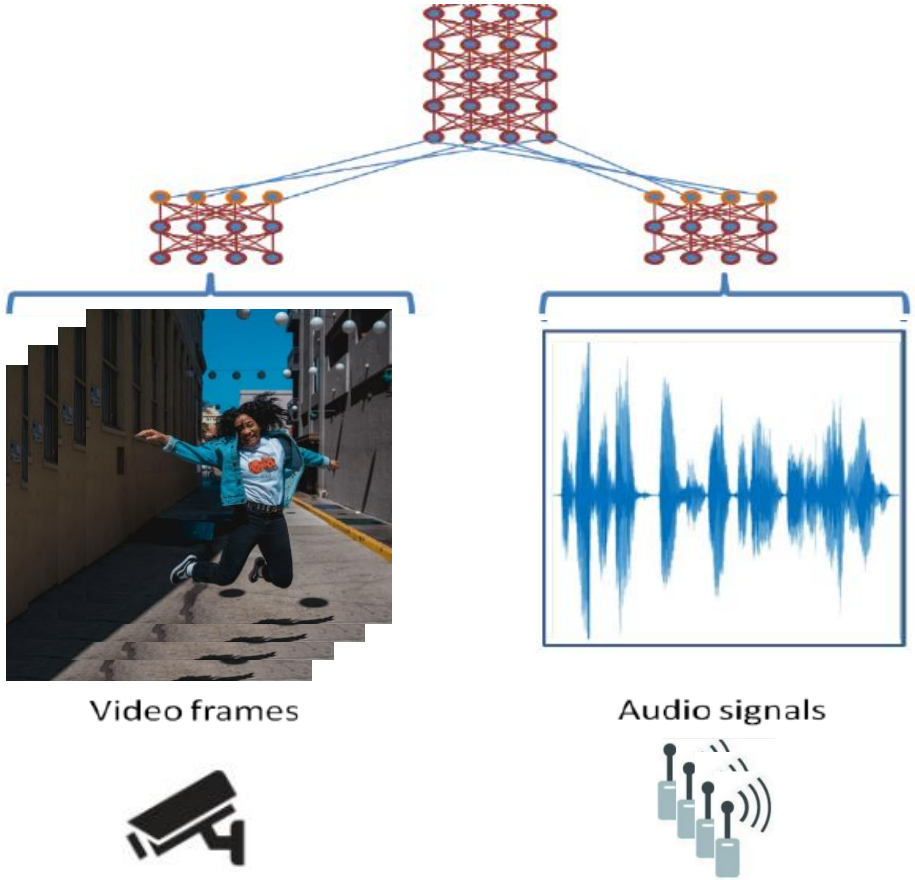
Safer roads for cyclists and pedestrians
Monitoring driver's behaviour
Monitoring traffic conditions/anomalous events
Data collection to support policy and infrastructure upgrades

Malta

Multimodal perception and intelligence

Project's rationale: *explore hidden correlations in synchronous streams of audio, visual and other data to increase classification accuracy of audio-visual/environmental events.*

- **Audio-visual analytics**



Early fusion – human – like perception!

Example: emotion detection

Multimodal perception and intelligence

Project's rationale: *explore hidden correlations in synchronous streams of audio, visual and other data to increase classification accuracy of audio-visual/environmental events.*

- **Anomaly detection in smart cities**

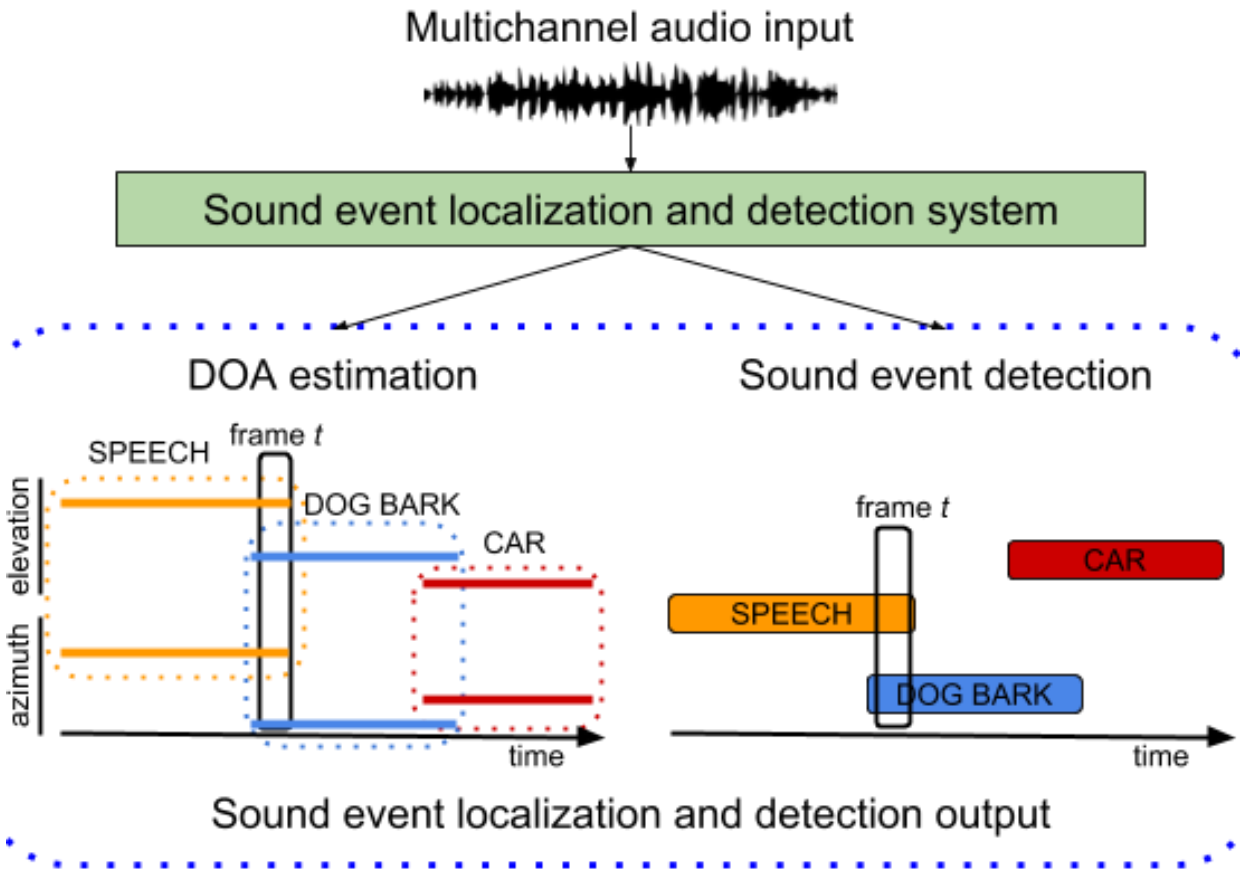
Detect the presence of an unexpected/unseen object, sound, action, etc.

Low visibility conditions: *exploit complementary **audio** information*!

Multimodal perception and intelligence

Project's rationale: *explore hidden correlations in synchronous streams of audio, visual and other data to increase classification accuracy of audio-visual/environmental events.*

- **Audio AI intelligence**

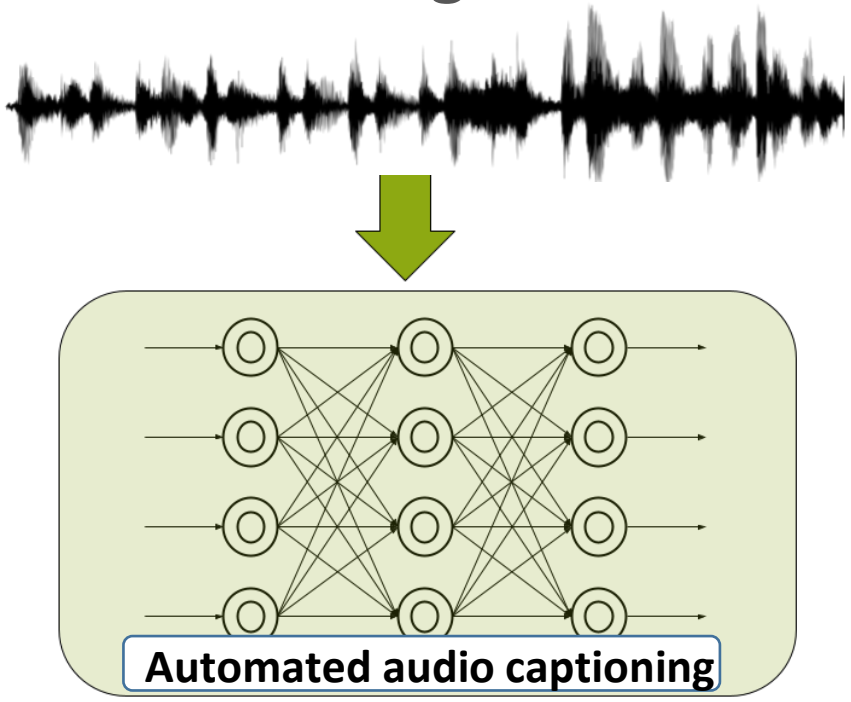


Acoustic analysis and mapping of city scenes!

Multimodal perception and intelligence

Project's rationale: *explore hidden correlations in synchronous streams of audio, visual and other data to increase classification accuracy of audio-visual/environmental events.*

- **Audio AI intelligence**



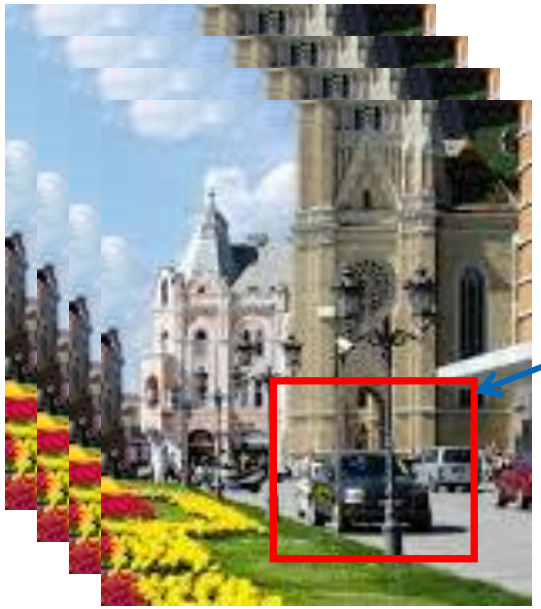
"Two people chatting on a busy street."

Sound-enabled environmental awareness

Multimodal perception and intelligence

Project's rationale: *explore hidden correlations in synchronous streams of audio, visual and other data to increase classification accuracy of audio-visual/environmental events.*

- **Advanced visualizations: Attention maps**

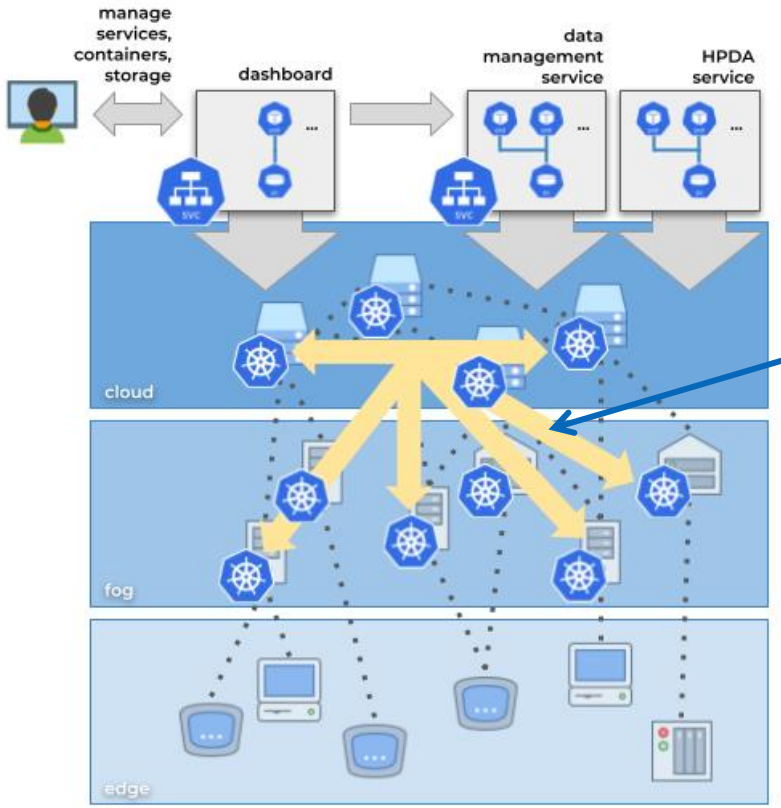


Effective real-time support for decision makers!

E2F2C Continuum Computing

Project's rationale: *capitalize on the vast amount of distributed resources in a Smart City infrastructure to achieve faster, better and deployment optimized analytics (faster time to decision, higher accuracy, less communications,…) .*

- **Optimal data and service placements**



Optimally distribute AI tasks!

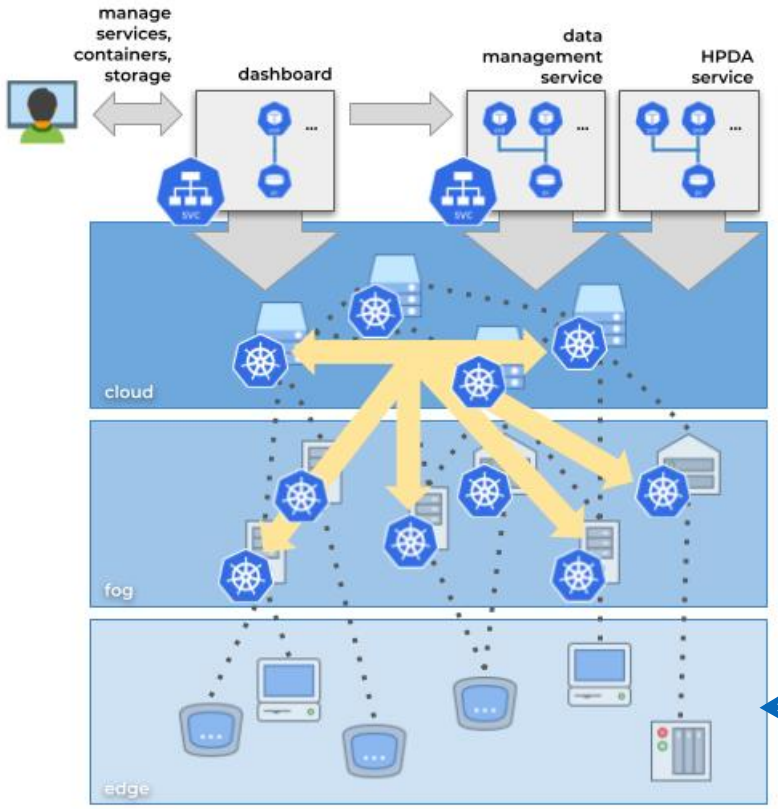
Bring the data where needed/best utilized by the tasks!

E2F2C Continuum Computing

Project's rationale: *capitalize on the vast amount of distributed resources in a Smart City infrastructure to achieve faster, better and deployment optimized analytics (faster time to decision, higher accuracy, less communications,…)*

- **Edge processing**

  - Model compression
  - Edge anonymization – only anonymized features sent to the cloud
  - Edge security
  - Edge GPU accel. – pattern matching

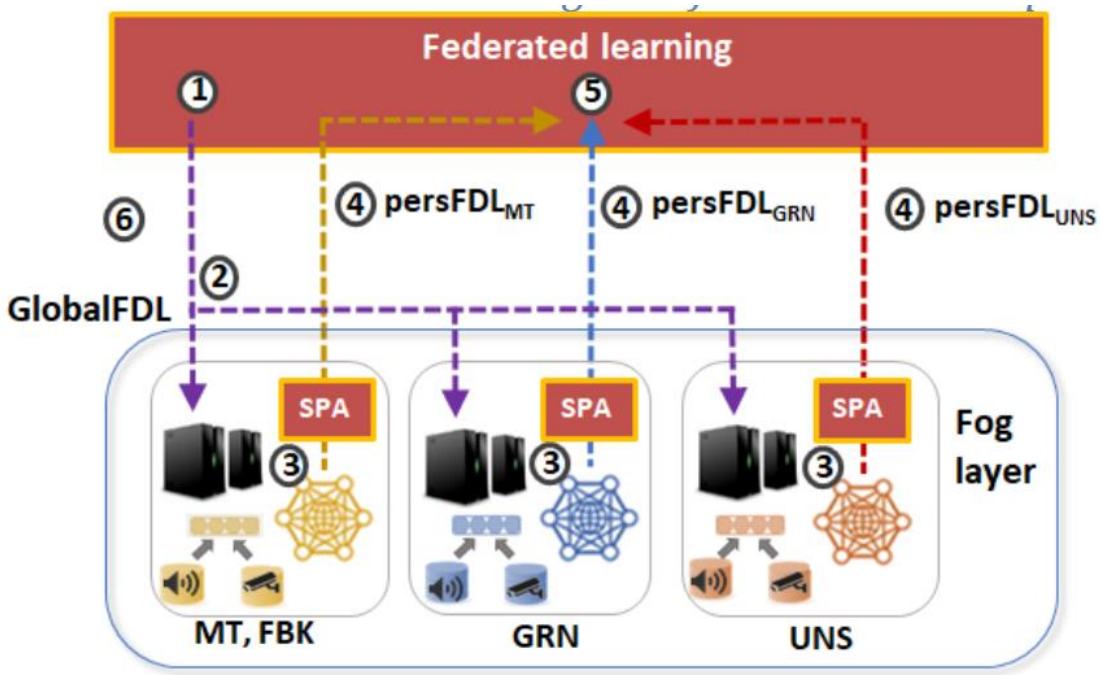


Most of computations will be performed here, close to the data sources!

E2F2C Continuum Computing

Project's rationale: *capitalize on the vast amount of distributed resources in a Smart City infrastructure to achieve faster, better and deployment optimized analytics (faster time to decision, better accuracy, less communications,…)*

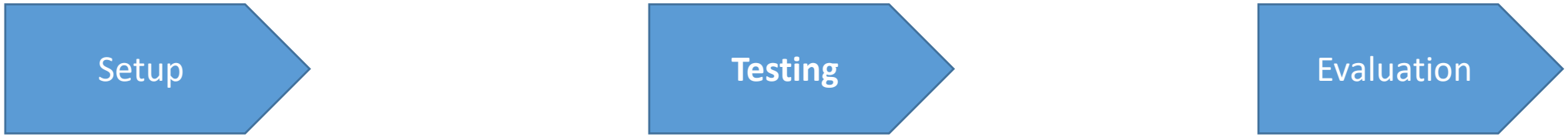
- **Federated learning (FL) framework**



Harnessing information from different datasets through **collaborative training of AI models** while **keeping the data private**

Benchmarking continuum

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



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

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

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.

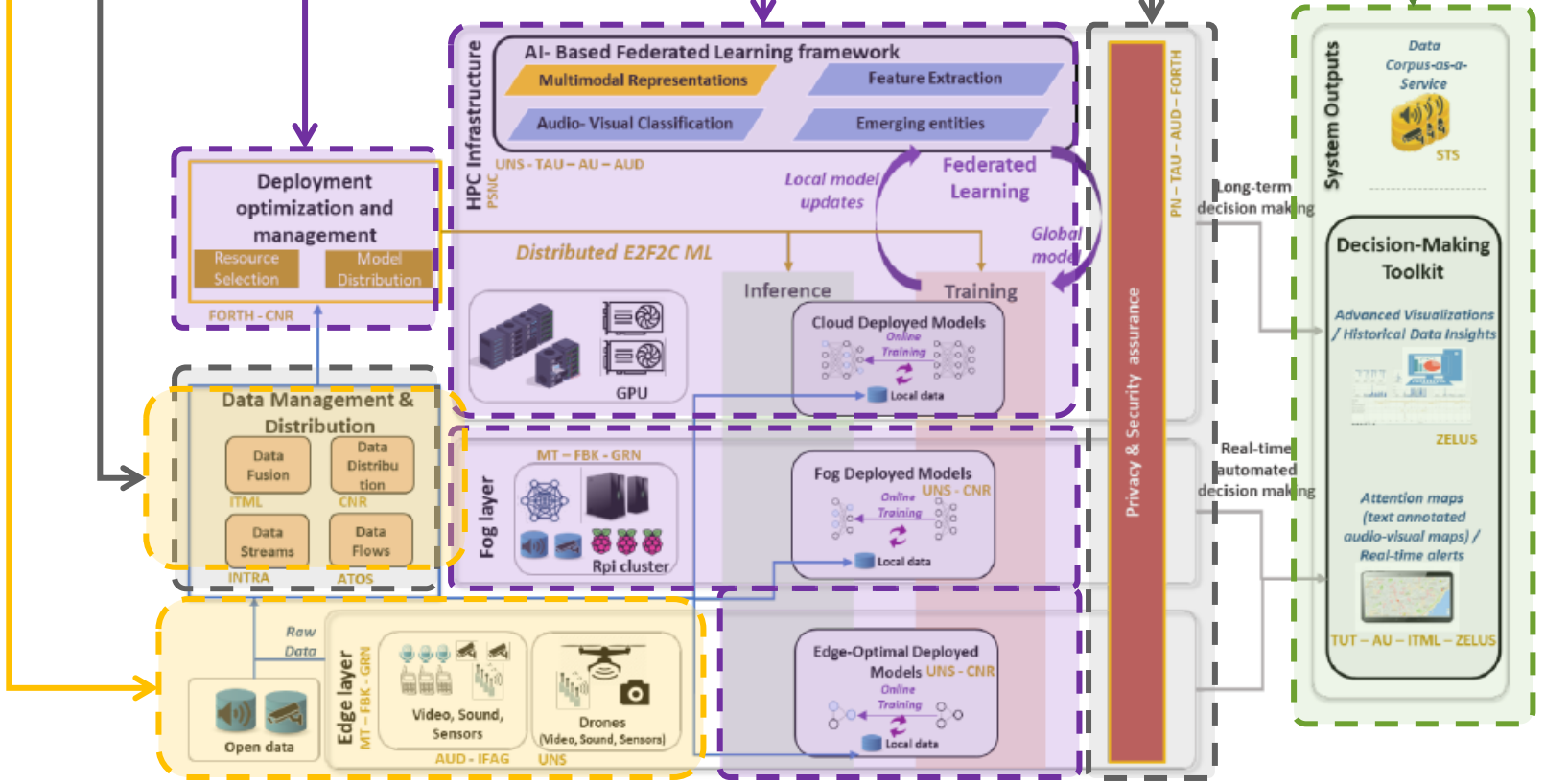
MARVEL Pipeline description

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)



MARVEL components and functionalities: Pipeline mapping

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)

Android app
sensMiner (audio streams, GPS, audio tags)
Devices (Data types)
GRNEdge
Cameras (video streams)
Advanced MEMS microph. (audio streams)
Drones
AI-enabled devices
SED@Edge kit
Data management toolkit

Data management toolkit
Data Fusion Bus
DatAna
Hierarchical Data Distrib.
StreamHandler
Annotation
iHEARu-PLAY platform
Privacy protection
AudioAnony
VideoAnony
VAD toolkit (devAIce)
Security assurance
EdgeSec

Multimodal AI
Audio-visual anomaly detect.
Audio-visual crowd counting
devAIce
Acoustic scene classification
Automated audio captioning
Sound event det. and local.

E2F2C computing framework
Karvdash deployment fram.
PerFeL federated learning
GPURegex acceleration
DynHP NN compression

Decision making toolkit
SmartViz
Attention maps
Audio-Visual summaries
Medium to long term business analytics

MARVEL Data Corpus as a Service
Advanced queries
Service level agreements