EUROPEAN BIG DATA VALUE FORUM

DATA NEXUS

SCAN HERE!

Innovative Approaches to Extreme Data

October 2, 14:00 - 15:30

EFRA EMERALDS EXA4MIND EXTRACT NEARDATA GRAPH MASSIVIZER

ORGANISED BY

BDV BIG DATA VALUE ASSOCIATION

IN COLLABORATION WITH

AI & AUT EXPO

IDEAL-IST

NTP Neumann Technológiai Platform

HUN REN SZTAKI

HORIZON-CL4-2022-DATA-01-05:

Extreme data mining, aggregation and analytics technologies and solutions

Provide better technologies, tools and solutions for **data mining** (searching and processing) of extreme data.

Extreme data is defined as data that exhibits one or more of the following characteristics, to an extent that makes current technologies fail: increasing **volume, speed, variety; complexity/diversity/multilinguality** of data; the **dispersed data sources; sparse/missing/insufficient** data/extreme variations in values).

The technologies and solutions are expected to discover and **distil meaningful, reliable and useful data** from heterogeneous and dispersed/scarce sources and deliver it to the requesting application/user **with minimal delay and in the appropriate format**.

EFRA

Extreme Food Risk Analytics

EFRA aims to develop the first analytics-enabled, green data space for AI-enabled food risk prevention.

The project explores how extreme data mining, aggregation and analysis could address major scientific, economic and societal challenges surrounding the safety and quality of the food that Europeans consume.

Funded by the European Union

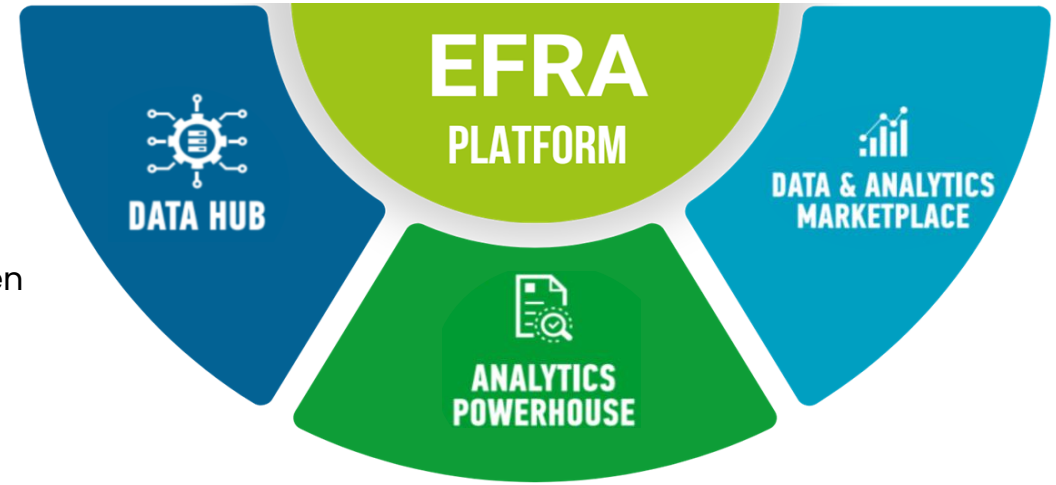
SCAN ME

EFRA

Extreme Food Risk Analytics

EFRA
EXTREME FOOD RISK ANALYTICS

Data Hub
Search, mine & process dispersed, multilingual, heterogeneous, deep/hidden food safety data sources



Data & Analytics Marketplace
Discover, purchase/use and contribute raw data, AI models, and analytics services

Analytics Powerhouse
Distill & combine useful insights from available data and train privacy-preserving, green, explainable risk prediction AI algorithms

DATA NEXUS
Funded by the European Union

WAGENINGEN UNIVERSITY & RESEARCH
EFRA
EMERALDS
EXA4MIND
EXTRACT
NEARDATA
GRAPH MASSIVIZER

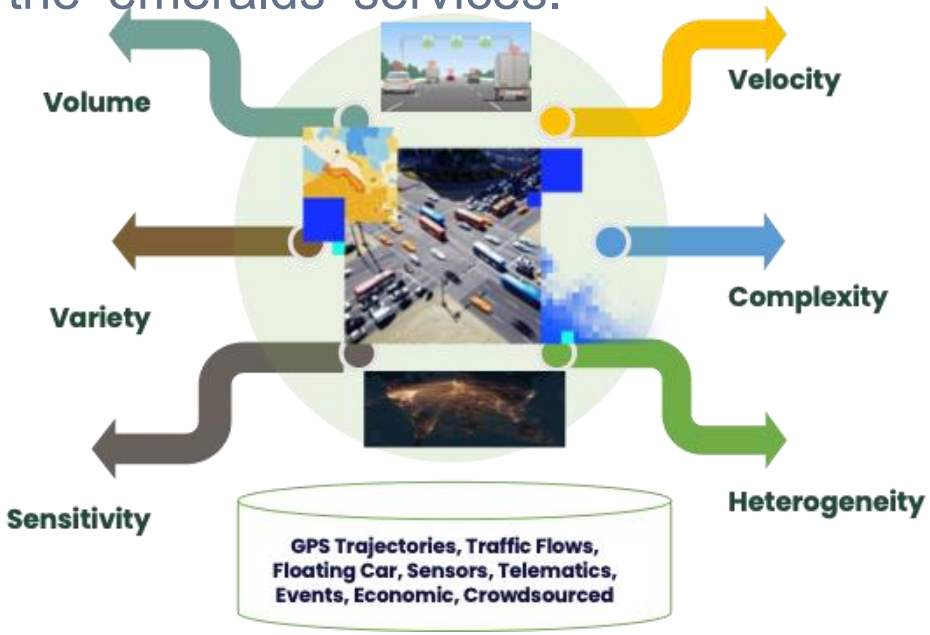
SCAN ME

EMERALDS

Extreme-scale Urban Mobility Data Analytics as a Service

The goal is to create an an urban data-focused Mobility Analytics-as-a-Service (MAaaS) efficient, interoperable and easy-to-deploy, consisting of the ‘emeralds’ services.

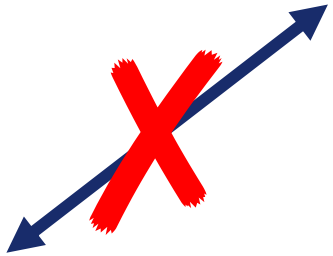
- harness **the untapped potential of extreme urban mobility data** through **efficient AI-driven systems**
- … thus, support industrial decision-making tasks at the appropriate level of the **Computing Continuum (CC)**

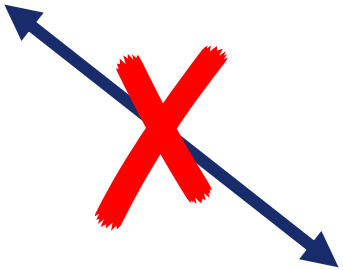


EXA4MIND: Motivation

Big Data Analytics –
Hot Topic Since Years

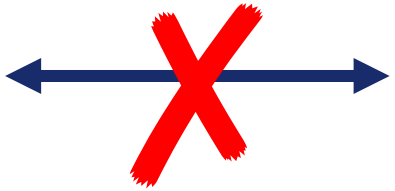
… yields great results with AI and database technology; **yields much more with adequate computing power.**





Excellent EU Supercomputing Facilities with HPC, Cloud, GPU and Quantum Systems

… do simulations and complex workflows but **not so much work on Databases & Extreme Data Analytics.**



European Data Spaces, EOSC and EUDAT

… do excellent research data management **without digging into** complex Big/Extreme Data.

EXA4MIND: Extreme Data Analytics…

… using EU Supercomputing & Data Ecosystems

EXTRACT

A distributed data-mining software platform for EXTReme dAta Across the Compute conTinuum

Integrating cloud, edge, and HPC technologies for trustworthy, accurate, fair and green data mining workflows for actionable knowledge

- Enable the development of **complex and secure data mining workflows**
- Develop **novel data-driven orchestration mechanisms** to efficiently deploy and execute data mining workflows
- Fully **exploit the performance capabilities of the compute continuum** to address extreme data characteristics
- Deliver the EXTRACT software platform and demonstrate its benefits in **2 use cases: Personalized Evacuation Routes** and **Transient Astrophysics**

Visit: https://extract-project.eu/ for more information

SCAN ME

EUROPEAN BIG DATA VALUE FORUM
DATA NEXUS
Funded by the European Union
EFRA EXTREME FOOD RISK ANALYTICS
EMERALDS
EXAHMIND
EXTRACT
NEARDATA

EXTRACT

A distributed data-mining software platform for EXTReme dAta Across the Compute conTinuum

A holistic approach to extreme data for social impact

The Personal Evacuation Routes (**PER**) use-case will serve to guide citizens in an urban environment (the city of Venice) through a safe route in real time.

Challenges are related to:

- the logistic of Venice (crowded attraction points, narrow streets, water canals)
- the difficulty to test the solution in real life (is not easy to start a real alarm)
- the management of the flow of extreme data in near real time
- Maintaining the security and the privacy of personal data

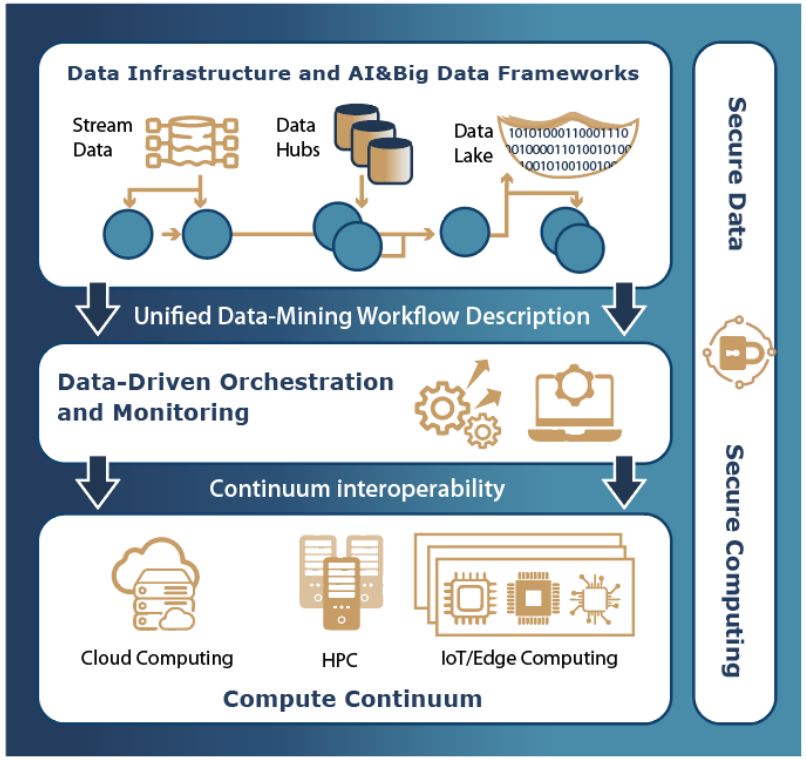
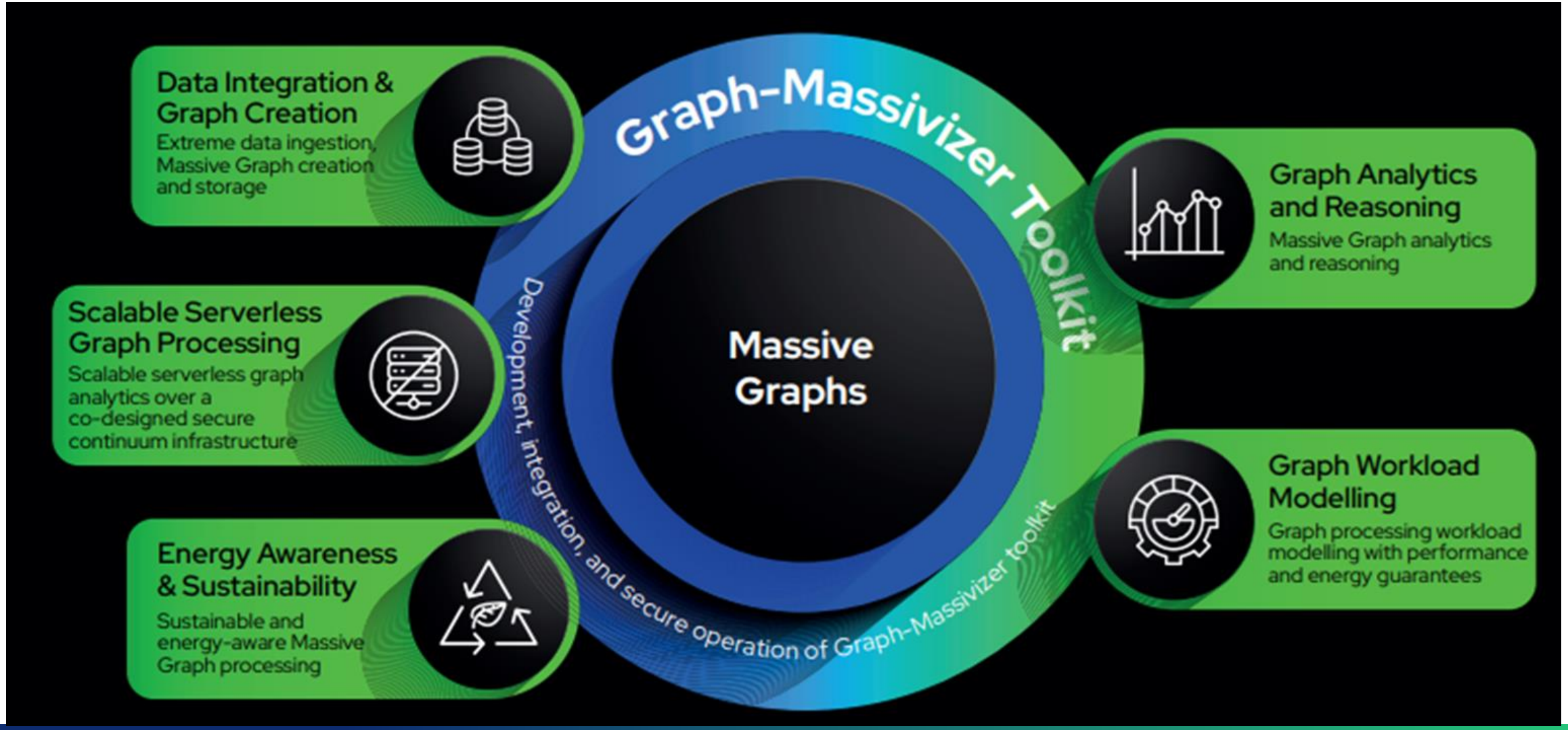


Figure 1. Main EXTRACT platform components . BSC©

Graph-Massivizer

Extreme and Sustainable Graph Processing for Urgent Societal Challenges in Europe



NEARDATA

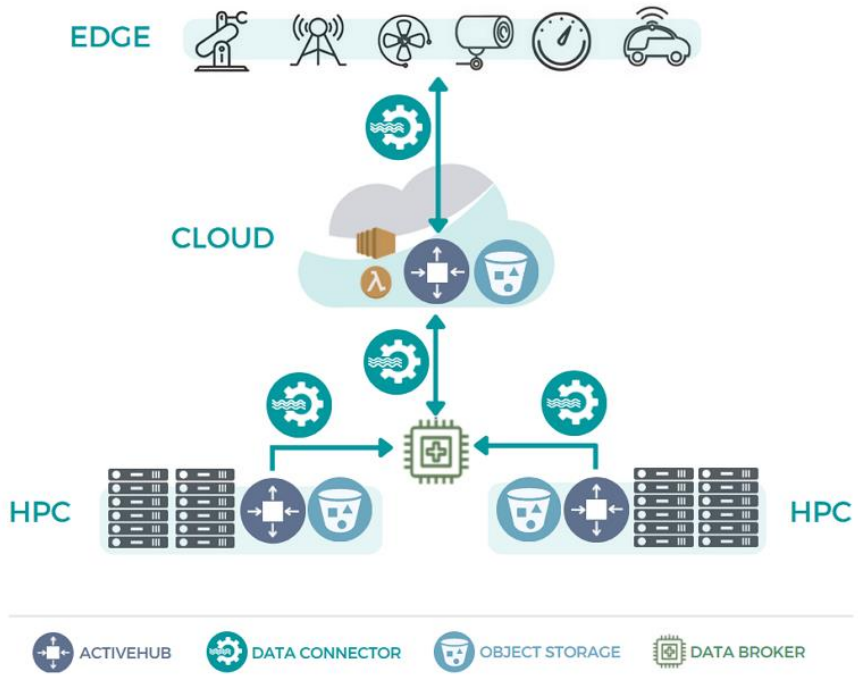
Extreme Near-Data Processing Platform

The goal of NEARDATA is to create an extreme data infrastructure mediating data flows between Object Storage and Data Analytics platforms across the Compute Continuum:

O-1 Provide high-performance near-data **serverless data connectors** that optimize data management operations to **efficiently present data to analytics platforms**.

O-2 Support **real-time video streams** but also event streams that must be **ingested and processed** very fast to Object Storage.

O-3 Create a Data Broker service enabling **trustworthy data sharing and confidential orchestration** of data pipelines.



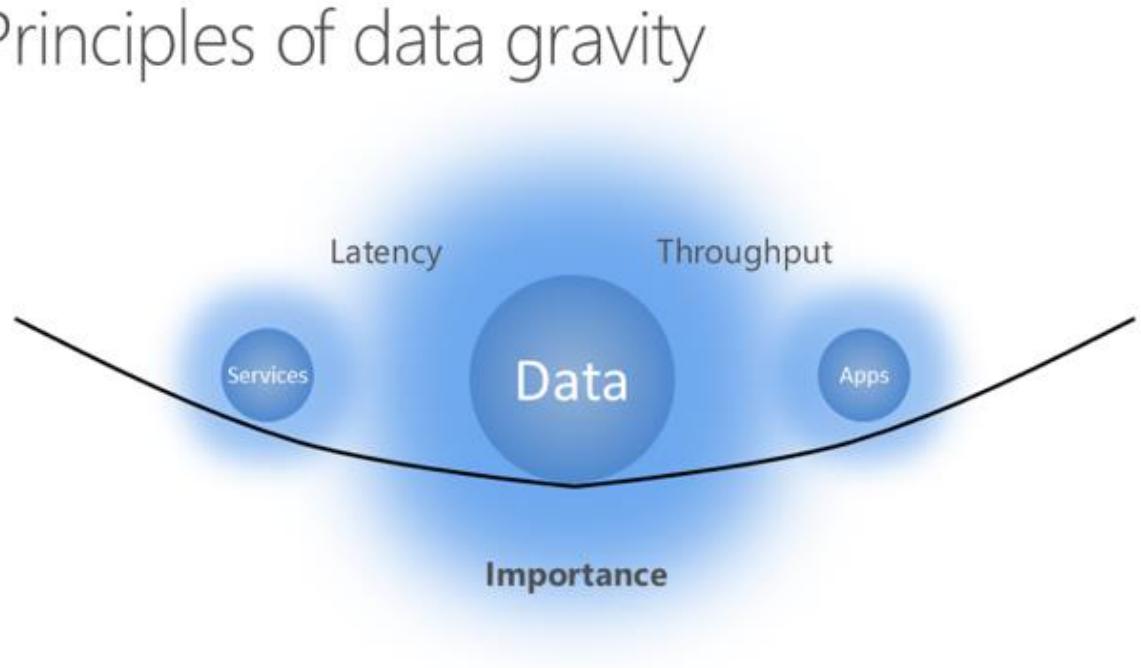
NEARDATA

Extreme Near-Data Processing Platform

NEARDATA

Why Locality ?

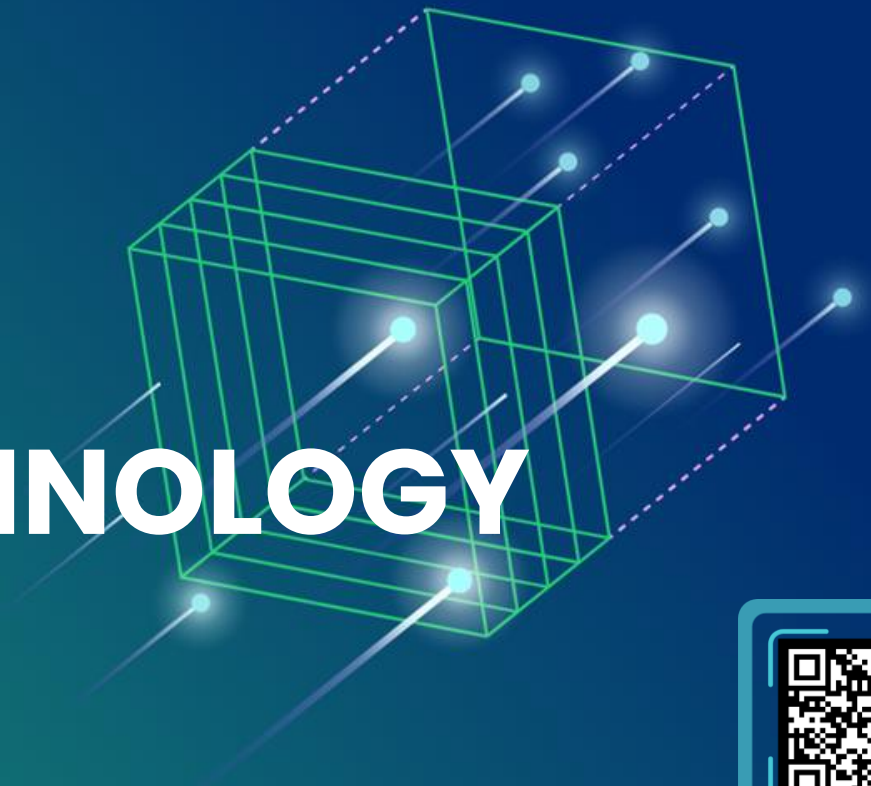
- Volume
- Privacy
- Low latency
- Hardware Acceleration



EUROPEAN BIG DATA VALUE FORUM

DATA NEXUS

PROJECT TECHNOLOGY & USE-CASES



SCAN HERE!

ORGANISED BY

BDV BIG DATA VALUE ASSOCIATION

IN COLLABORATION WITH

AI & AUT EXPO

IDEAL-IST

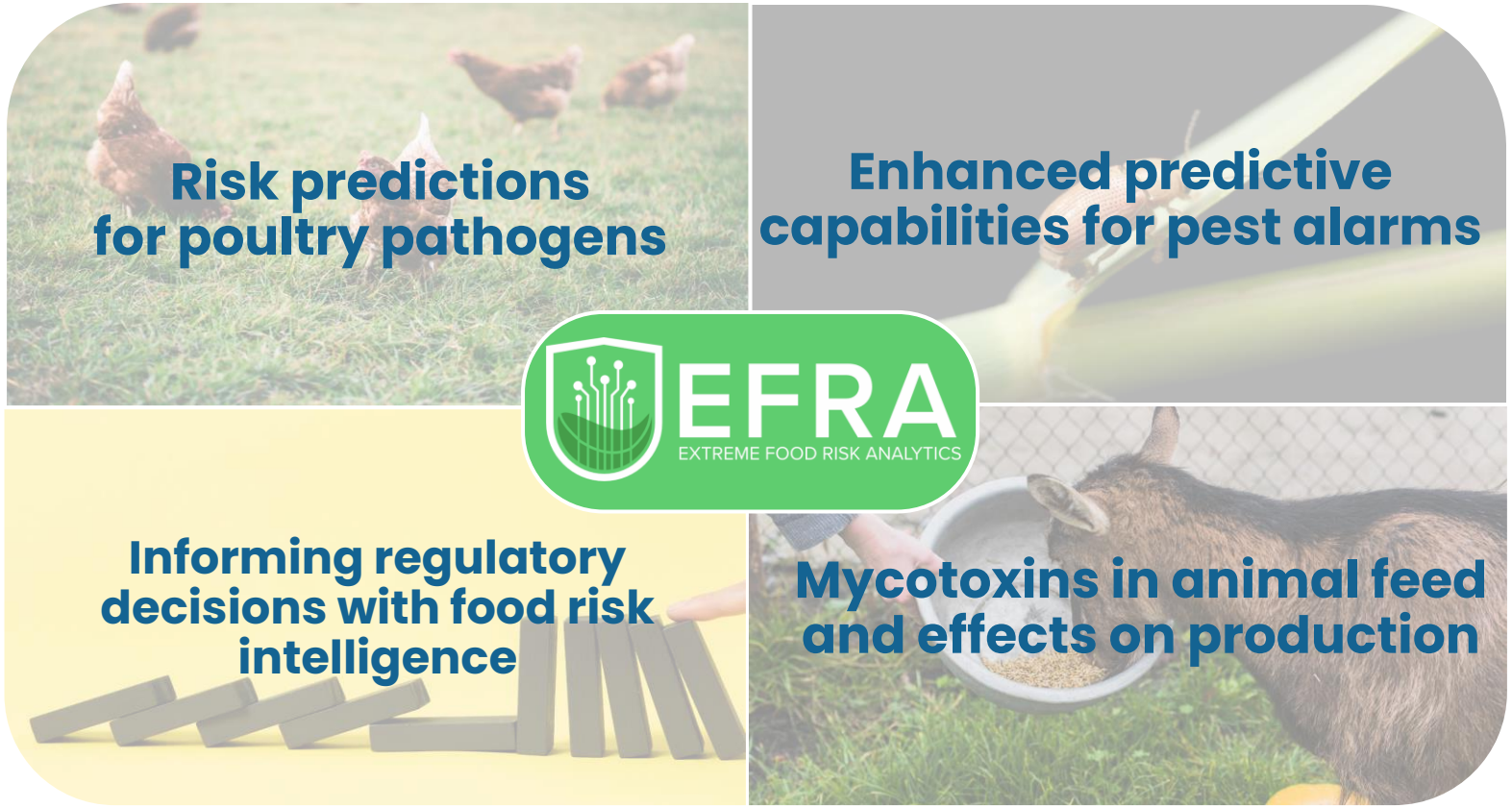
NTP Neumann Technológiai Platform

HUN REN | SZTAKI

EFRA

Extreme Food Risk Analytics

EFRA
EXTREME FOOD RISK ANALYTICS



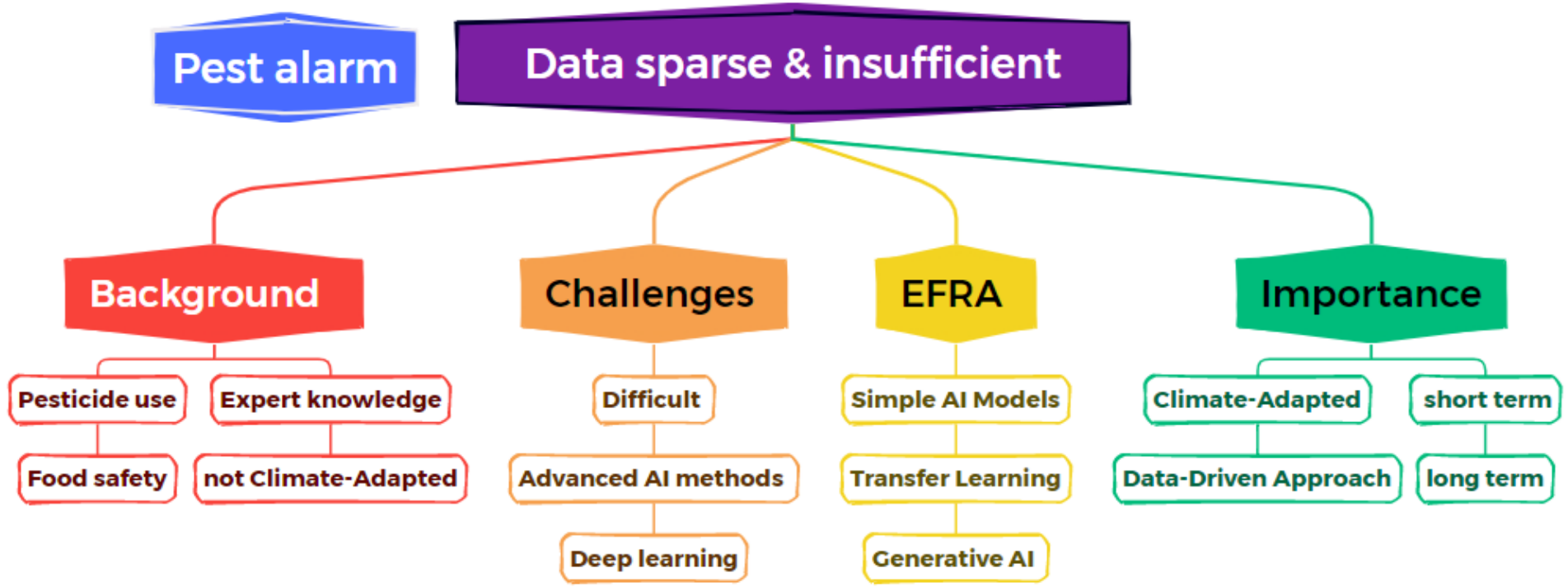
SCAN ME

EFRA
Extreme Food Risk Analytics

WAGENINGEN UNIVERSITY & RESEARCH

EFRA EXTREME FOOD RISK ANALYTICS

Case Study



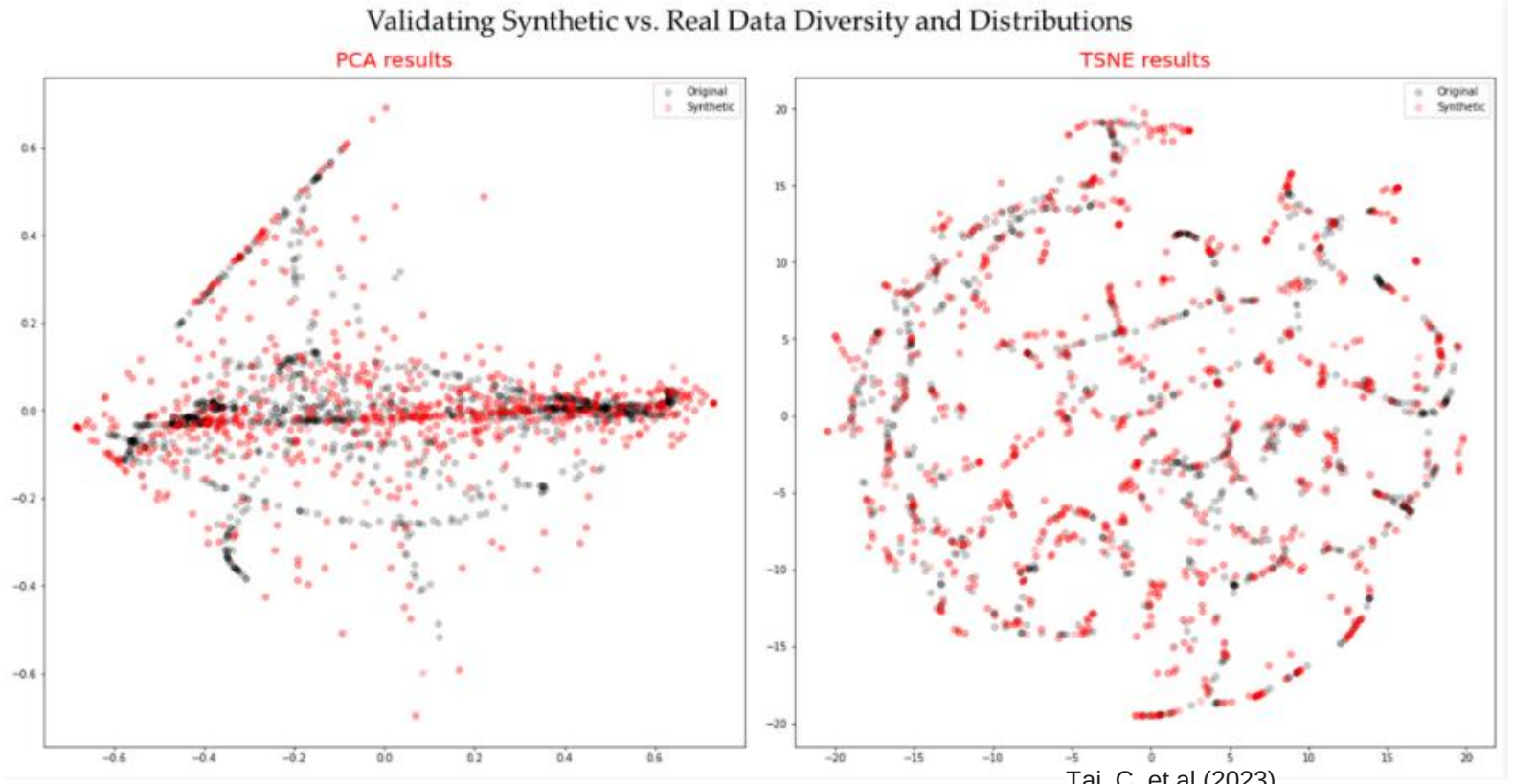
SCAN ME

EFRA
Extreme Food Risk Analytics

WAGENINGEN UNIVERSITY & RESEARCH

EFRA EXTREME FOOD RISK ANALYTICS

Generative AI | Synthesize multivariate agricultural data



Tai, C. et al (2023)

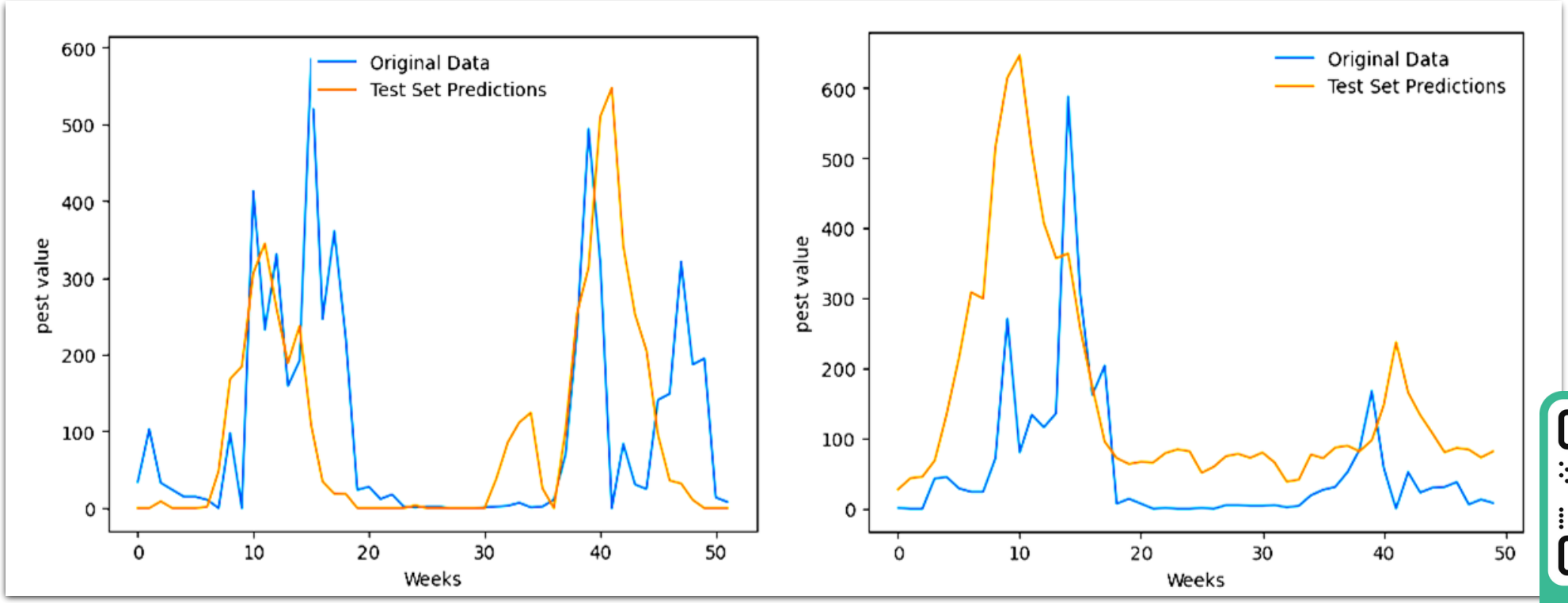
SCAN ME

EFRA
Extreme Food Risk Analytics

WAGENINGEN UNIVERSITY & RESEARCH

EFRA
EXTREME FOOD RISK ANALYTICS

Transformer | Pest disease prediction in rice

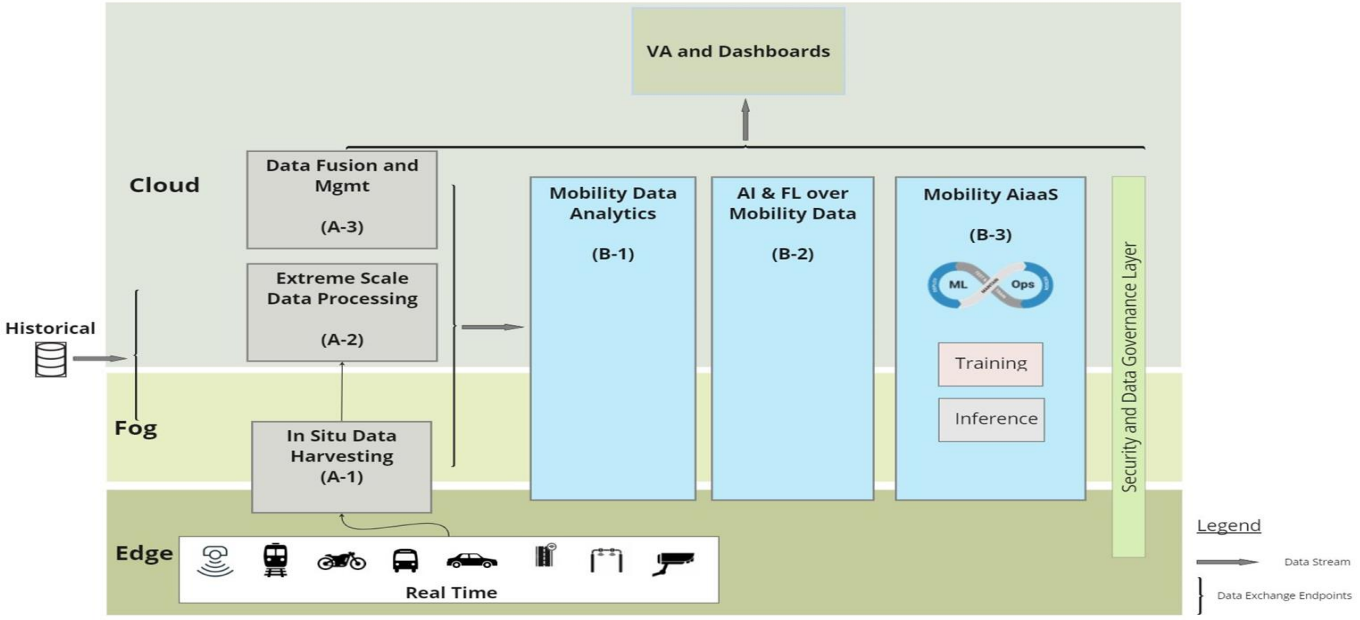


SCAN ME

EMERALDS

Extreme-scale Urban Mobility Data Analytics as a Service

- an **AIaaS-oriented reference architecture for mobility data**
- a suite of **extreme-scale data processing & analytics** methods



Data Processing and Analytics Services
Extreme-scale Orchestrator & Privacy Evaluator
Extreme-scale Map Matching & Weather Integrator
Spatio-Textual Range & top-K Queries
Mobility Data Compression and Fusion
Trajectory, Floating Car Data and Travel Time Analysis
Probabilistic Trip Chaining and Drop Off Prediction
Crowd Density & Parking Garage Occupancy Prediction
Traffic Flow / State Prediction

Data Infrastructure and Development Frameworks
Containerized EMERALDS Toolset & Hosting Services
CARTO Analytics Toolbox
Data Management Component & Data Transfer Protocols
Secure Communication & Network Intrusion Detection System
ML Models Development & Deployment Framework

EMERALDS

Extreme-scale Urban Mobility Data Analytics as a Service

The Hague

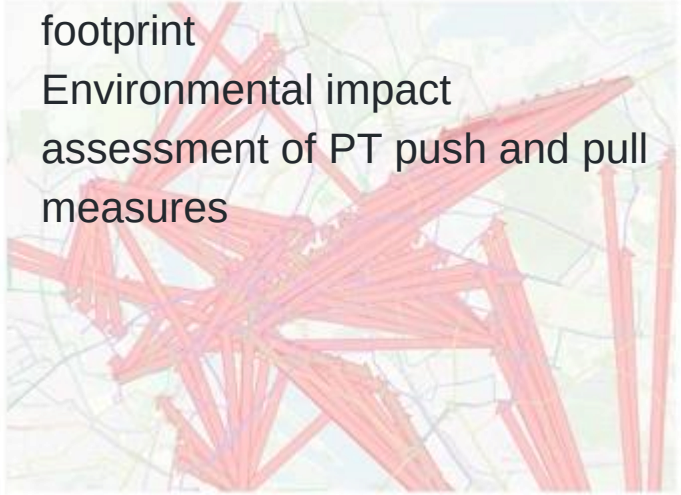
- Improved safety and security measures during events
- Evidence-based policy design
- Suggestions on crowd management and risk measures
- Reduced risks related to overcrowding and emergencies

Rotterdam

- Reduced traffic congestion and improved traffic flow
- Real-time traffic management and optimization. (forecast-as-a-service providers)
- Timely actuation of traffic management systems

Riga

- Efficient public transport services with optimized routes/schedules
- Calculation of the total system cost and passenger carbon footprint
- Environmental impact assessment of PT push and pull measures



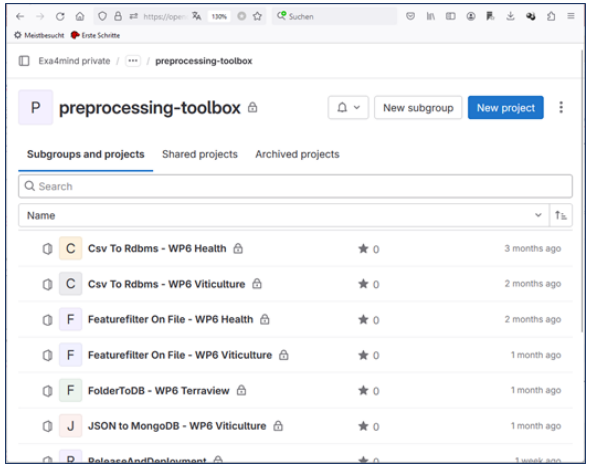
EXA4MIND: Portable Platform

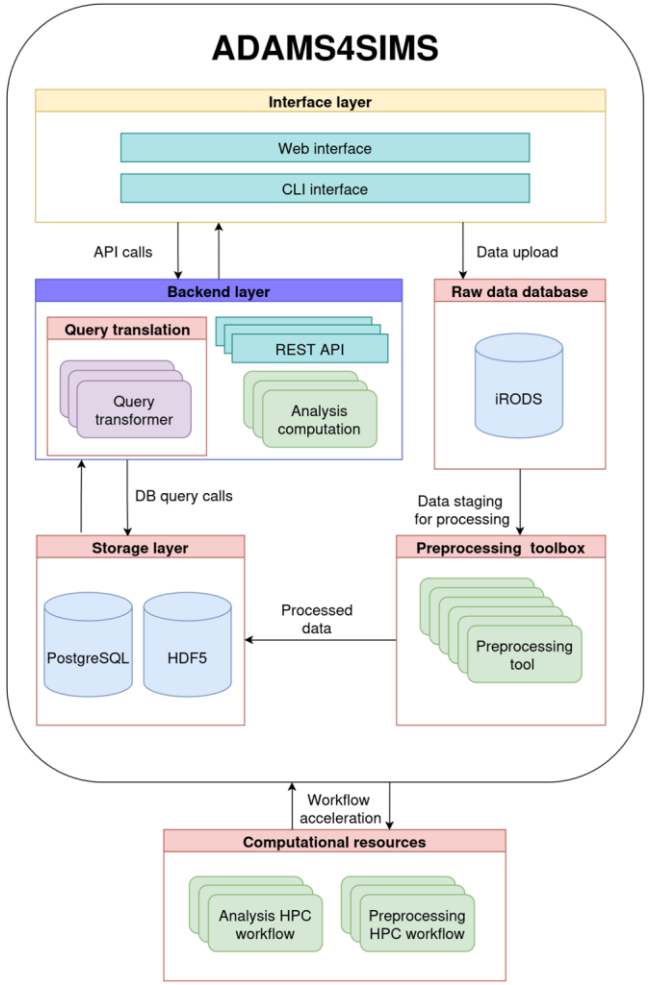
For info visit https://exa4mind.eu

- **Easy-instantiation modular** platform
- Brings together Extreme Data Analytics, Supercomputing and Data Spaces
  - Cross-backend querying and **natural-language queries**
  - Standardised **preprocessing**
  - Efficient **Data Transfer**
  - **Streaming/Indexing/ Caching**
- **Automatized Data-Driven Workflows**

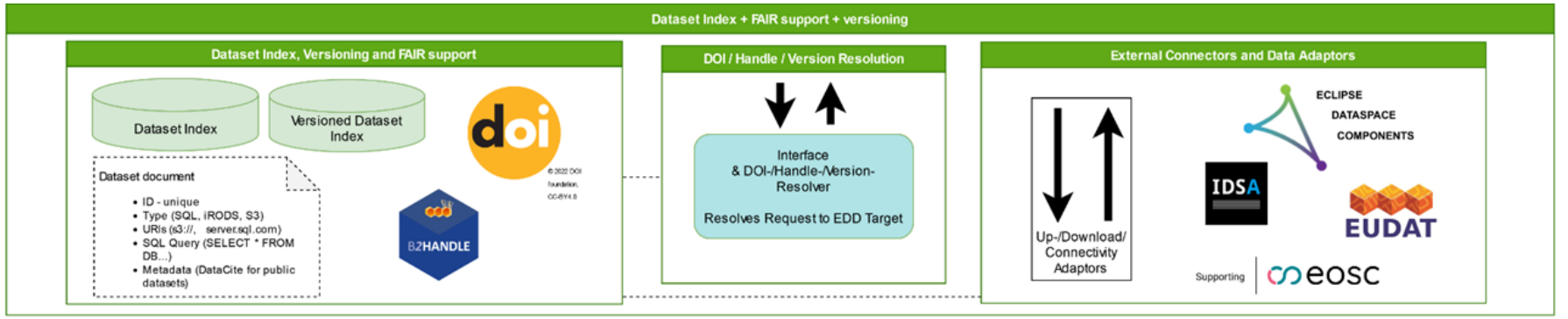
Image: CC-BY; vectorportal.com

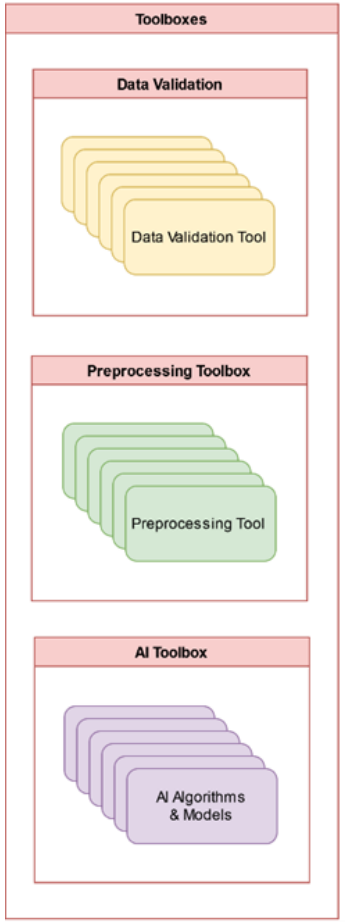
E.g. MD simulation data-mining site relying on selected modules

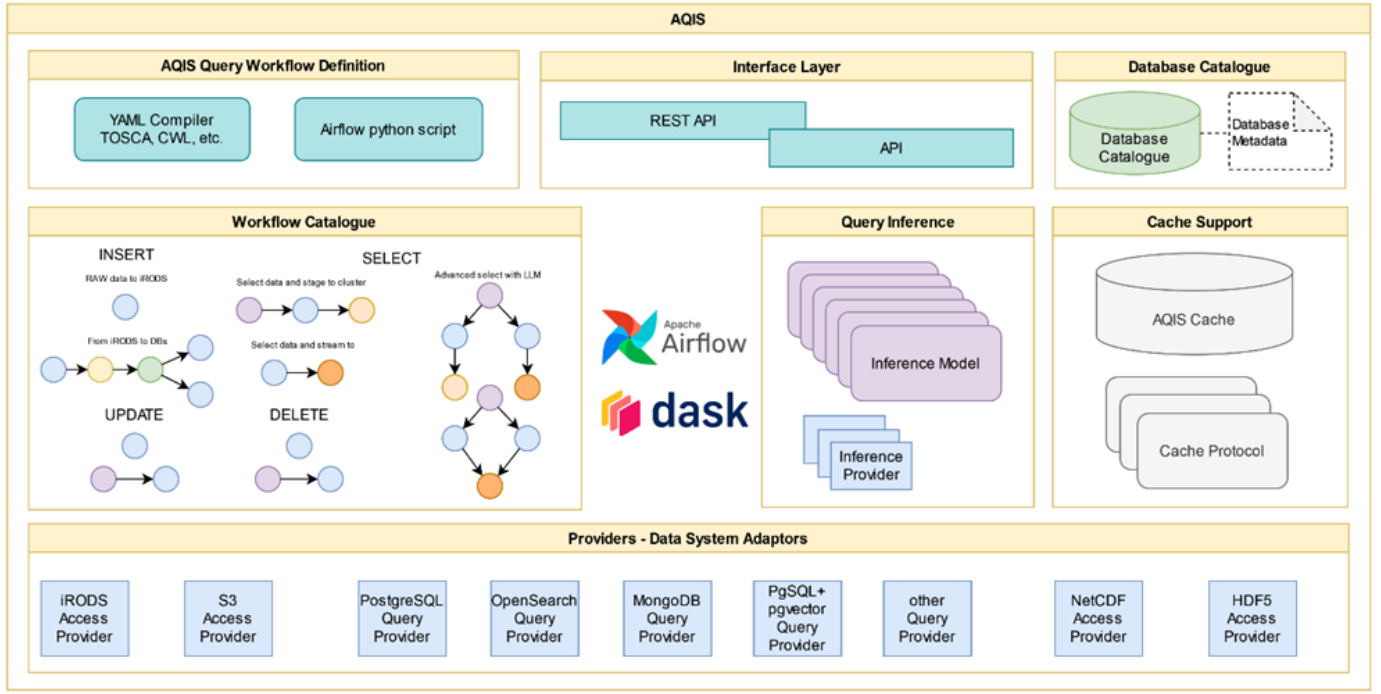


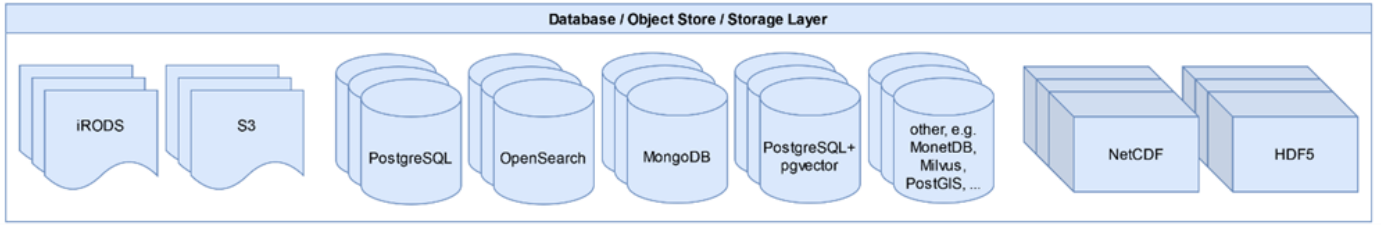


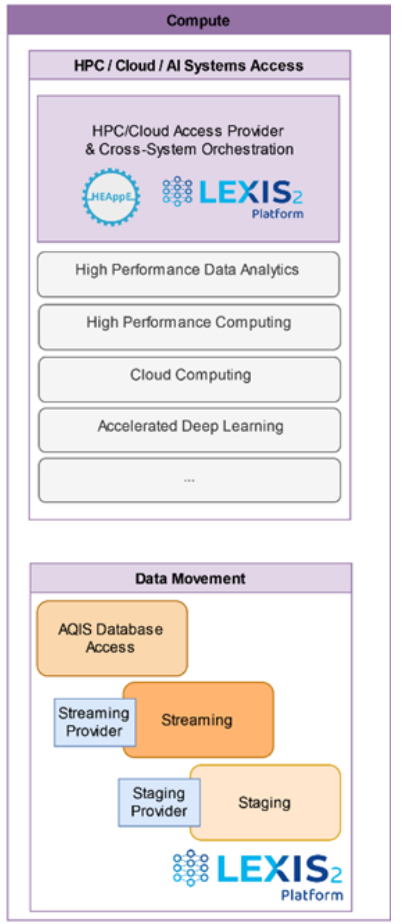
EXA4MIND











SCAN ME

EXA4MIND: Application Cases

SCIENTIFIC – Molecular simulations
WP4

ADAMS4SIMS, an innovative Automated Data Mining System for Systematic Improvement of Molecular Simulations accuracy and predictability will combine services and on-/offline data sources.

Data: 0.1-1 PB range

INDUSTRY – Automotive Car Sensor Recordings
WP5

Advanced Driver Assistance Systems (ADAS) need Extreme Data annotation, search and analytics. Methods, AI models and framework are co-developed in EXA4MIND.

Data: 1-100 PB range

SME – Smart farming / viticulture
WP6

Easy and quick ingestion of large amounts of agriculture focused data, as well as efficient and space-saving re-distribution of computed/ingested (meta-)data.

Data: 0.01-1 PB range

SME – Health
WP6

Demonstration of viability of Extreme Data Database (EDD) for Extreme Data processing use cases with health and societal datasets, based on direct user interaction.

Data: 0.01-0.1 PB range

SCAN ME

EXTRACT

A distributed data-mining software platform for EXTReme dAta Across the Compute conTinuum

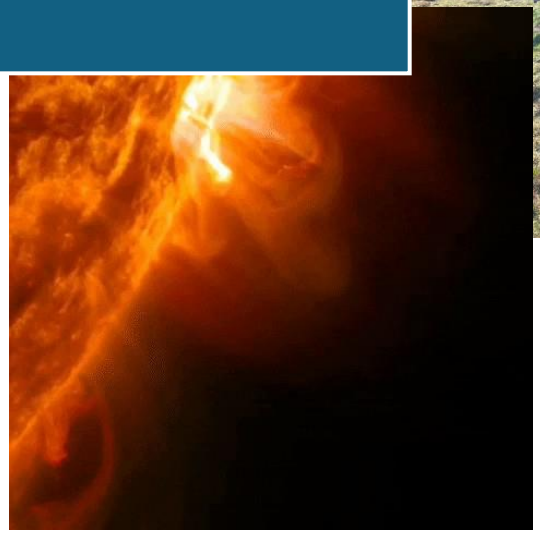
PER

Personalised Evacuation Route in Venice

SCAN ME

TASKA

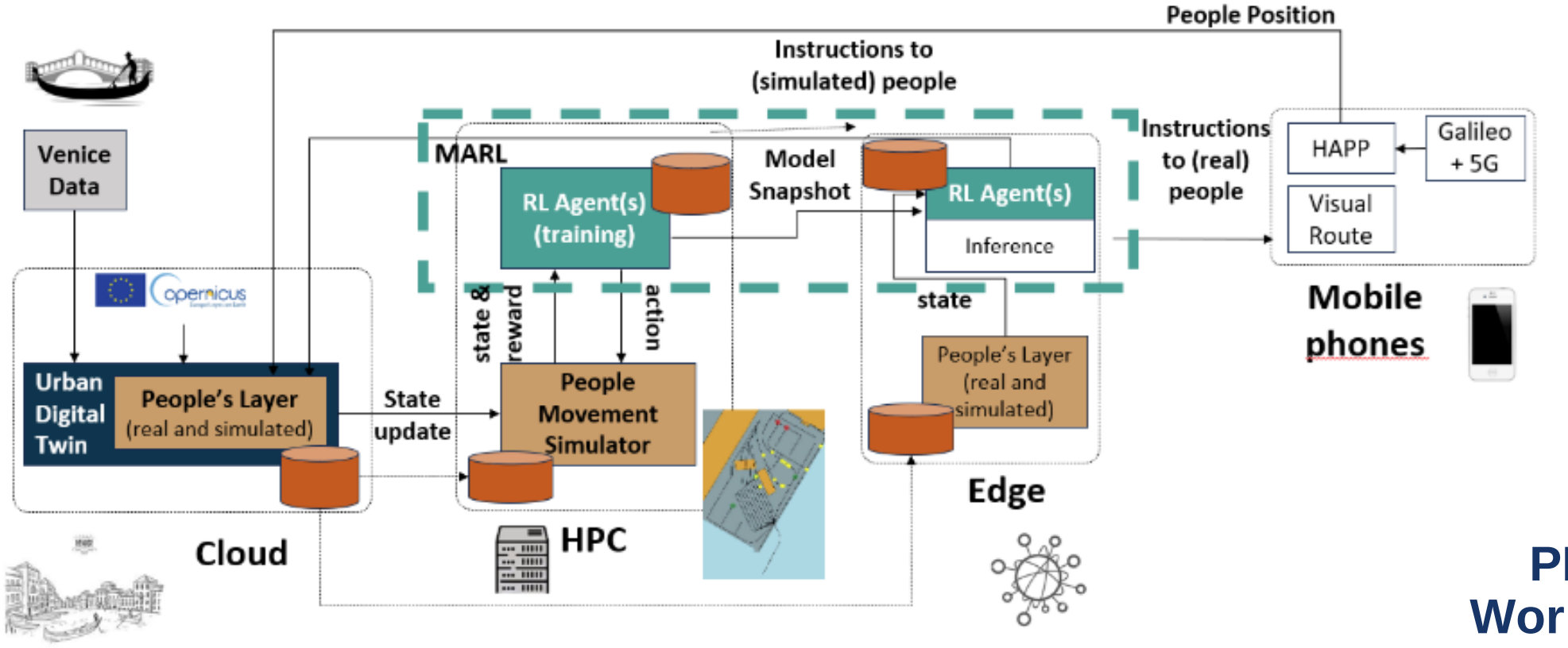
Transient Astrophysics with a square Kilometre Array Pathfinder



SCAN ME

EXTRACT

A distributed data-mining software platform for EXTReme dAta Across the Compute conTinuum



PER Workflow

EXTRACT

A distributed data-mining software platform for EXTReme dAta Across the Compute conTinuum

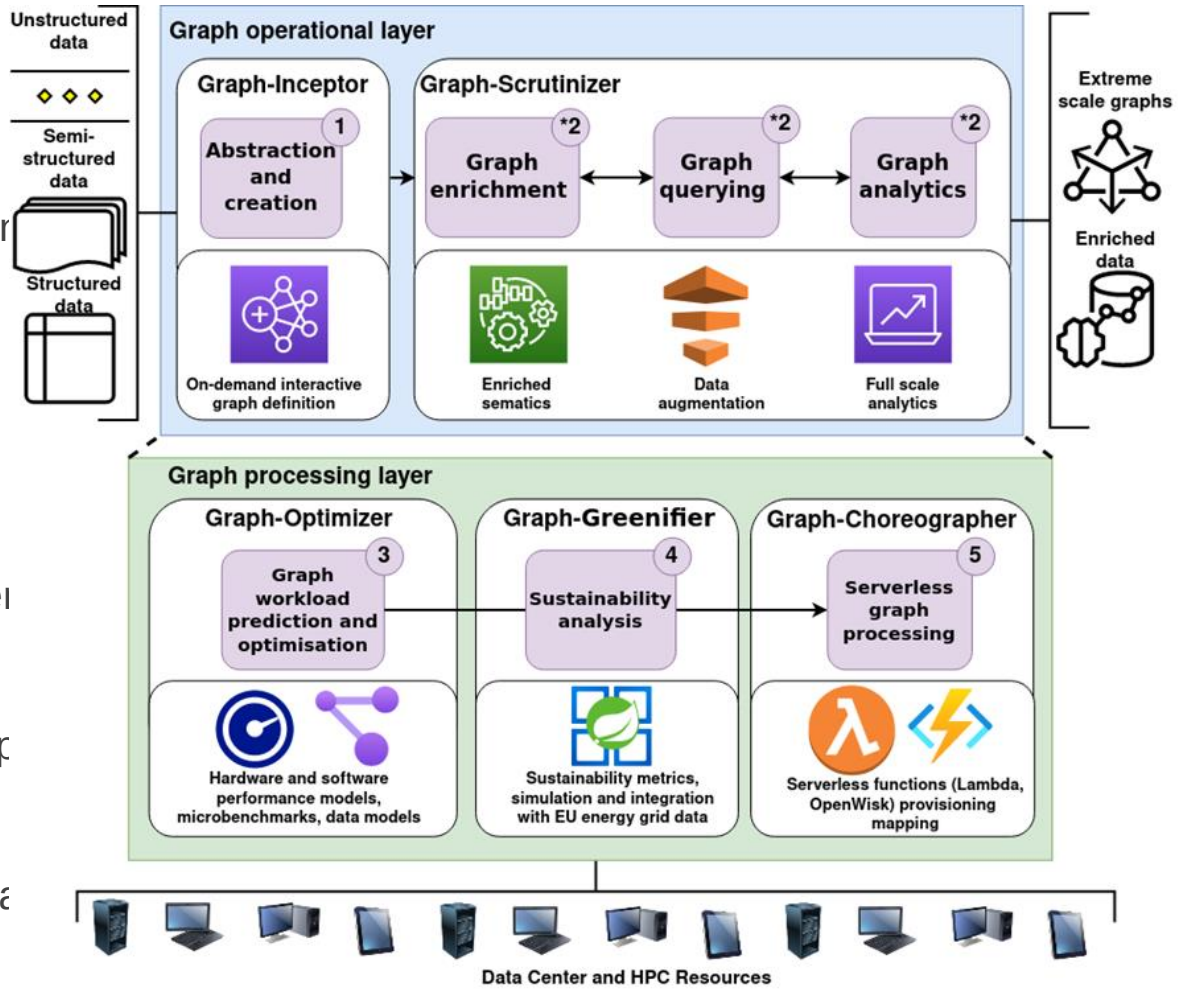
Once the **Urban Digital Twin** has conveyed all the data of interest to the **MARL**, it is able to generate a trained model on the current state of the city of Venice.

The trained model is then inferred for each individual simulated user, in order to obtain the **Personalized Evacuation Routes** valid for that specific situation.

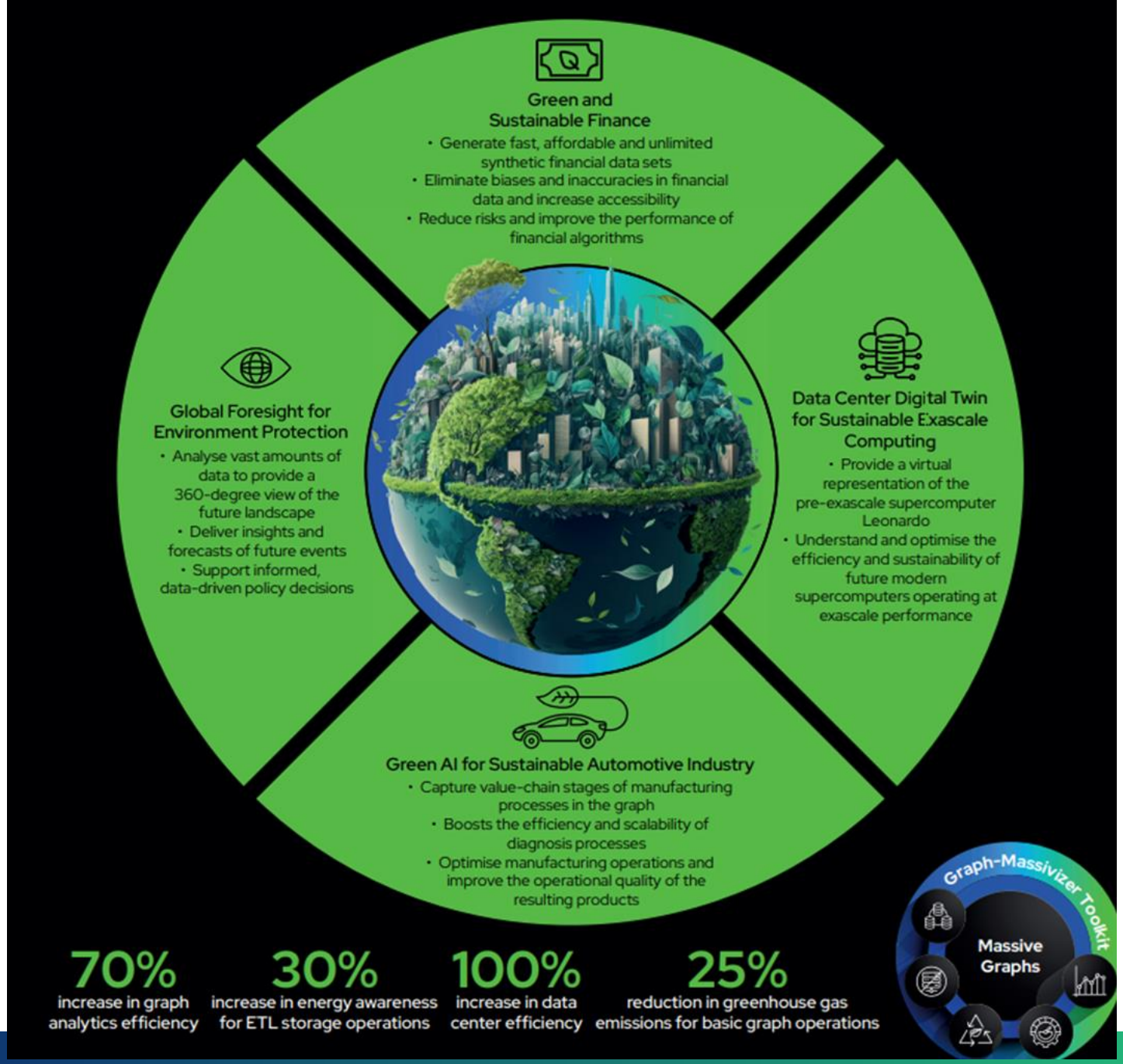
The preliminary results of all the operations developed by the **PER** use case can be viewed in the animation: the users, following the generated indications, head towards the safety points (**in green**) escaping the danger points (**in red**)

Graph-Massivizer

- **Graph operational layer**
  - Graph-Inceptor: extreme data ingestion, massive graph creation storage
  - Graph-Scrutinizer: massive graph analytics and reasoning
- **Graph processing layer**
  - **Graph-Optimizer:** workload modelling with performance and e guarantees
  - **Graph-Greenifier**: sustainable and energy-aware massive grap processing
  - **Graph-Choreographer**: scalable serverless massive graph ana over the computing continuum



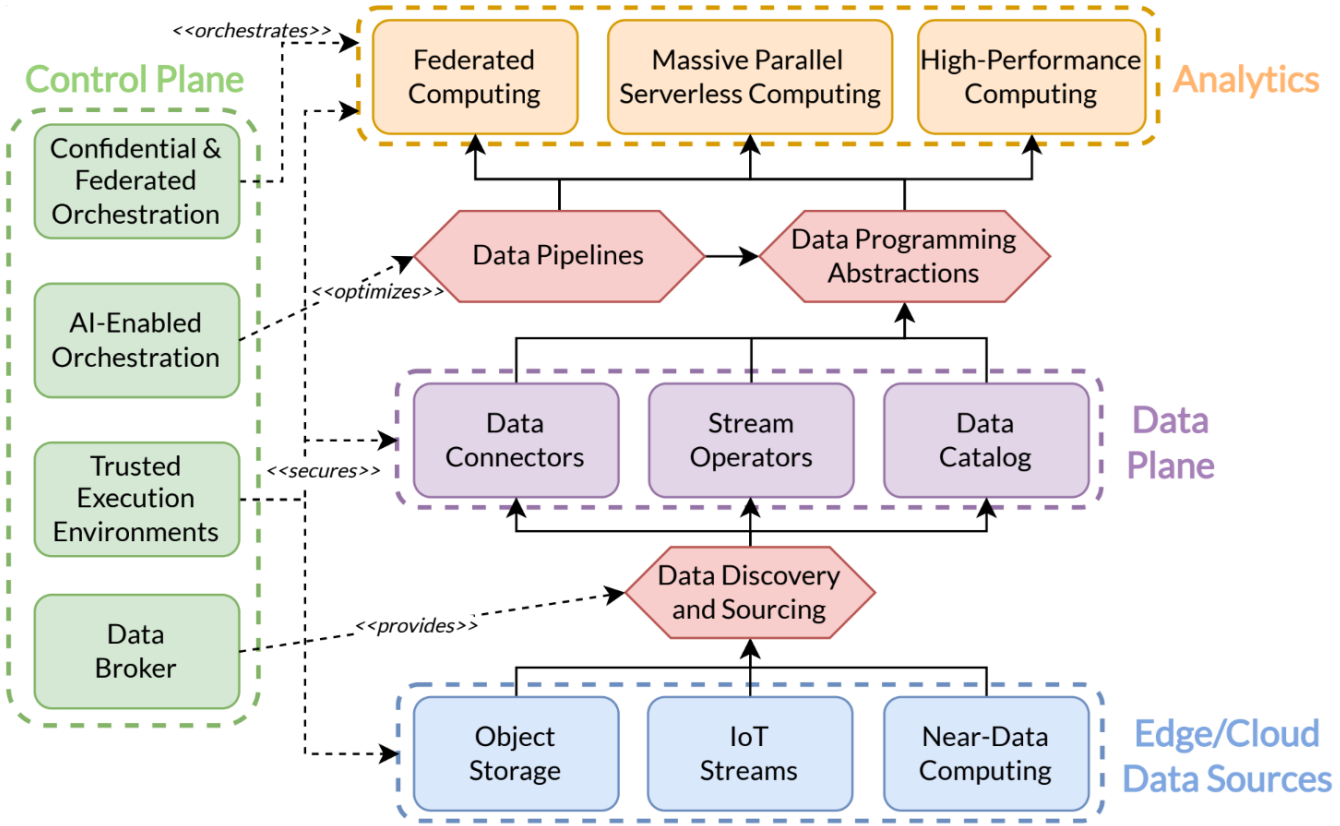
GRAPH MASSIVIZER



NEARDATA

Extreme Near-Data Processing Platform

NEARDATA Architecture



LITHOPS

METASPACE

Pravega
Reliable streaming storage

SCONE

NEARDATA

Extreme Near-Data Processing Platform

CHALLENGES:
Volume
Complexity

METABOLOMICS
NEARDATA cloud platform
for spatial metabolomics
[Cloud - Edge]

SURGERY
Real time surgery
video analytics
[Edge]

NEARDATA

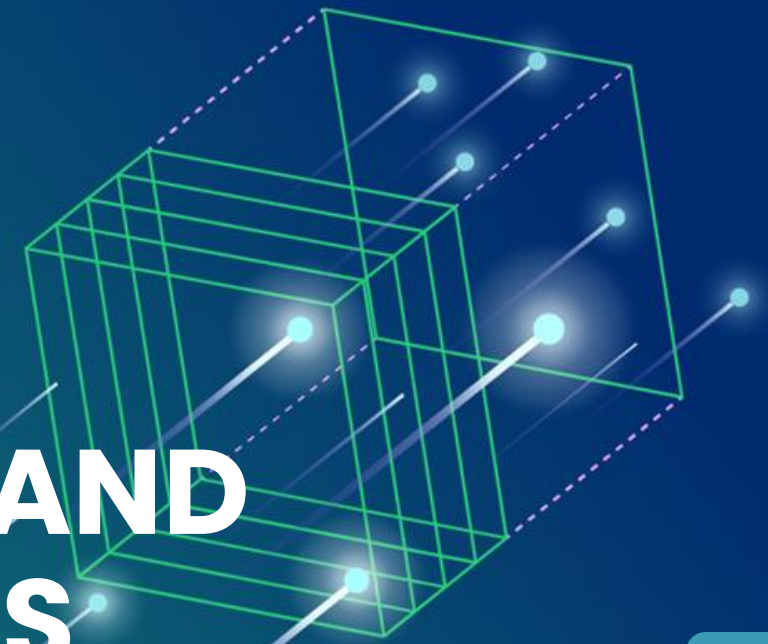
GENOMICS
Optimizing large genomic pipelines
[Cloud - HPC]

CHALLENGES:
Speed
Confidential

CHALLENGES:
Volume
Confidential

EUROPEAN BIG DATA VALUE FORUM

DATA NEXUS



DISCUSSION AND QUESTIONS

SCAN HERE!

ORGANISED BY

BDV BIG DATA VALUE ASSOCIATION

IN COLLABORATION WITH

AI & AUT EXPO

IDEAL-IST

NTP Neumann Technológiai Platform

HUN REN | SZTAKI

EUROPEAN BIG DATA VALUE FORUM

DATA NEXUS

THANK YOU!

SCAN HERE!

EFRA EMERALDS EXA4MIND EXTRACT NEARDATA GRAPH MASSIVIZER

ORGANISED BY

BDV BIG DATA VALUE ASSOCIATION

IN COLLABORATION WITH

AI & AUT EXPO

IDEAL-IST

NTP Neumann Technológiai Platform

HUN REN SZTAKI