EXAHMIND

EXtreme Analytics For MINing Data spaces

Jan Martinovič (IT4I, Coordinator)

Funded by
the European Union

EXA4MIND: Who we are

- **9** partners, **6** countries, **36** months
- EU contribution: **€ 4.9M**
- Overall budget: **€ 6M**
- **Coordinator**

  Jan Martinovič (IT4I)
- **Science & Co-Design Coordinator**

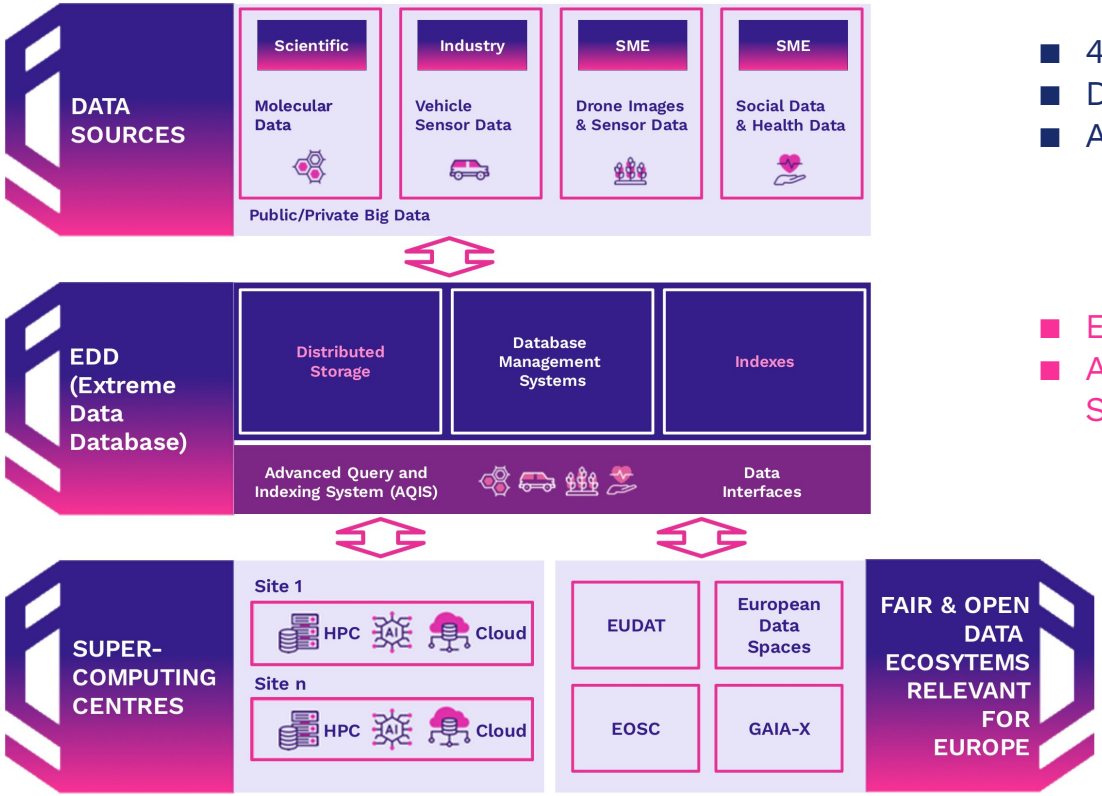
  Stephan Hachinger (LRZ)

EXA4MIND Technological Objectives

Integration for a leap forward in Extreme Data processing & valorisation

- Leverage the **Strongest European Computing Systems** for Extreme Data Mining and Analytics
- Develop a Unified Extreme Data **“Advanced Query and Indexing System”** and Data Analytics Capabilities Leveraging **High-Performance Data Libraries**
- Integrate High-Performance Databases with **European Data Spaces and FAIR** Distributed Data Management Approaches
- Development of **Fair, Trustworthy AI and Data Ecosystems** in the Context of Extreme Data Processing

EXA4MIND: What we do



- 4 Application Cases
- Data harmonisation
- Analytics

- Extreme Data Database (EDD)
- Advanced Query and Indexing System (AQIS)

- Staging and workflow automation
- Data access & publication

Application Cases

SCIENTIFIC – Molecular simulations

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 Sensors Recordings

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 – AGRICULTURE – Smart farming/ viticulture

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

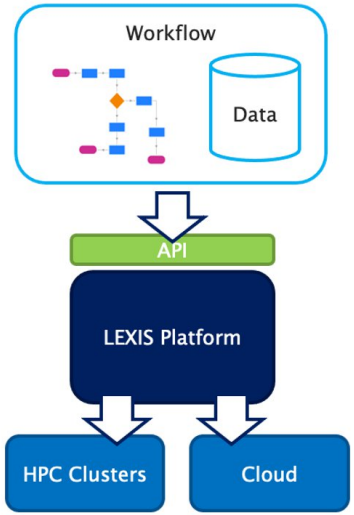
SME – Health

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

Key points of the LEXIS Platform

- **Complex computing workflows across HPC and Cloud**
  - Orchestration in geographical federation with *Atos Ystia, Apache Airflow and HEAppE*
  - Secure access to HPC resources through REST API
- **Distributed Data Management**
  - Federated data transfers, staging and discovery with *iRODS*
  - Data encryption and compression module
  - PID and replication features by EUDAT services
- **Web portal and interfaces for workflow set-up / execution**
  - Unified access to all services via OpenID-based (Keycloak) LEXIS AAI
  - Zero trust security model implemented
  - Already tested with EUDAT B2ACCESS and ready for other ID providers
- **Easy HPC/Cloud access including SMEs/Industry**
  - HPC-as-a-Service approach – REST API for all features of the platform
  - Accounting multiple HPC centres - *CYCLOPS and HEAppE*
  - Integration with *Puhuri* for resources management and additional accounting



Open Source

https://docs.lexis.tech

LEXIS and EXA4MIND Platforms Integration Plan

LEXIS THE PLATFORM

EXA4MIND

User interfaces
Web GUI
REST APIs

Resource management

Workflow orchestration

Distributed Data Infrastructure

HPC Parallel FS

POSIX

S3 Objects

CACHE

Indexes

SQL

iRODS

AAI & User management

Monitoring

Smart scheduling

Metadata & control APIs

Data staging and transfer for Extreme Data

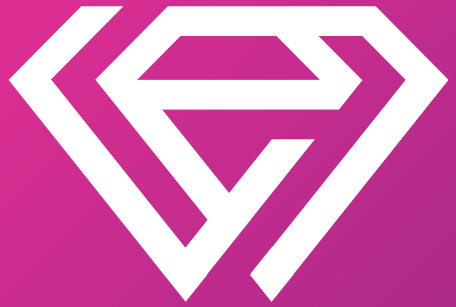
Supercomputing Centre Resources

Cloud compute

HPC Clusters

Remote storage

https://exa4mind.eu



Thank you – Contact Us!

Jan Martinovič

jan.martinovic@vsb.cz

Funded by
the European Union

Keep in touch & follow us: **@exa4mind** **/exa4mind**