DATAWEEK24
JOIN.LEARN.SHARE.GET VALUE

DATA NEXUS

Innovative Approaches to Extreme Data Challenges: DataNexus Cluster

Luxembourg, 10 December

Nuria de Lama, Consulting Director, IDC

BDV BIG DATA VALUE ASSOCIATION

IN COLLABORATION WITH

LNDS LUXEMBOURG NATIONAL DATA SERVICE

LUXEMBOURG INSTITUTE OF SCIENCE AND TECHNOLOGY LIST

Netcompany

DataNexus: Why?
Challenges

DATA NEXUS

- **Extreme data** that requires advanced techniques, tools, or infrastructure to process, analyze, and manage effectively. It leads to **innovations in cloud computing, machine learning, and distributed systems** to handle its challenges
  - Size, complexity, velocity, variety, volume at scale, sensitivity, extreme environments…
- Technical Challenges

Storage/ Scalability	**Data processing**	Data Integration	**Data quality/ Governance**	Real time Analytics
Privacy and Security	Cost (infrastructure)	**Data Analytics**	Latency (Computing Continuum)	**Data lifecycle management**

DataNexus: Why? Opportunities for the EU

DATA NEXUS

- **Strengths**
  - Research & Innovation ecosystems
  - Data-related regulations and governance
  - Competitiveness in sectors characterized by extreme data: manufacturing, healthcare, Earth observation
  - Open data and generally data availability
  - Interoperability and federation (e.g. GAIA-X)
- **Weaknesses/dependencies**
  - Cloud providers from outside EU → data sovereignty
  - Fragmented data infrastructures
  - From innovation to market → scalability
  - High competition by global analytics players
  - Talent retention

- Investment in data sovereignty and infrastructure
- Leveraging ethical AI and analytics
- Promote regional collaboration to foster cross-border cases (and avoid fragmentation)
- Focus on strategic sectors
- Cooperation between Public and Private sectors

DataNexus: what?

DATA NEXUS

Introducing the EU Cluster Harnessing Extreme Data using Advanced European Technologies

Seven EU projects collaborate around shared mission to extract meaningful insights from extreme data

Date 19/06/2024

As part of the Horizon Europe funding programme, seven pioneering projects have received funding to address the challenges posed by extreme data generated by IoT, industrial, business, administration, environmental, scientific and societal sources. These projects are pursuing innovative approaches that integrate cutting-edge technologies, including Artificial Intelligence (AI), Internet of Things (IoT), High-Performance Computing (HPC) and Edge/Fog/Cloud Computing.

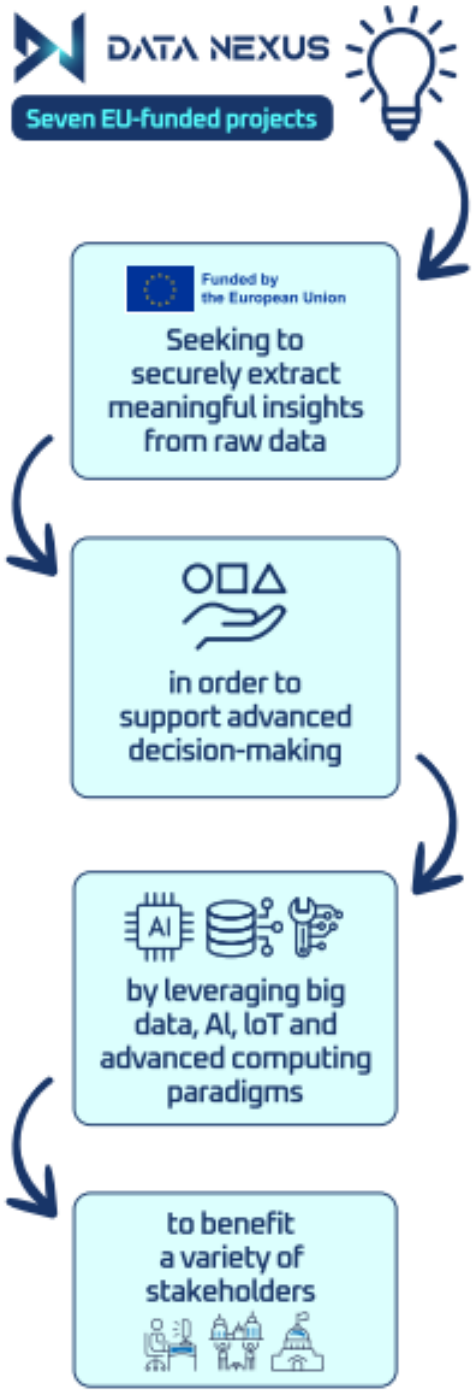
Under the 2022 Horizon Europe call for Digital and Space, topic "World Leading Data and Computing Technologies", the EU recognises the intrinsic value of data for EU's competitiveness. Projects funded under the cluster HORIZON-CL4-2022-DATA-01-05 - Extreme data mining, aggregation and analytics technologies and solutions seek to securely extract meaningful insights from raw data, facilitating advanced decision-making strategies.

Methodology and Novelty

Through unique methodologies and application cases, each of the seven projects contribute to strengthening European capacities for agile responses to urgent needs by developing and enabling the uptake of the next generation of computing and data technologies and infrastructures. Operating within the framework of European values, the projects emphasise human-centred and ethical development of digital and industrial technologies, promoting trustworthy technology that empowers end-users.

Funded by the European Union

- Projects funded under the cluster **HORIZON-CL4-2022-DATA-01-05 - Extreme data mining, aggregation and analytics technologies and solutions** seek to securely extract meaningful insights from raw data, facilitating advanced decision-making strategies.



DataNexus: Who?

DATA NEXUS

Graph-Massivizer – Researches and develops a high-performance, scalable, and sustainable platform for information processing and reasoning based on the massive graph representation of extreme data. **graph-massivizer.eu**

EMERALDS – Design, develop and create an urban data-oriented Mobility Analytics as a Service (MAaaS) toolset to exploit the untapped potential of extreme urban mobility data and as a result, improve urban mobility decision making. **emeralds-horizon.eu**

EXTRACT – Delivering a data-driven open-source platform integrating cloud, edge and HPC technologies for trustworthy, accurate, fair and green data mining workflows for high-quality actionable knowledge. **extract-project.eu**

SYCLOPS – Advancing AI/data mining for extremely large and diverse data for Europe and beyond, by democratizing its acceleration through open standards and a healthy, competitive, and innovating ecosystem. **www.syclops.org**

NEARDATA – Create an extreme data infrastructure mediating data flows between Object Storage and Data Analytics platforms across the Compute Continuum. **neardata.eu**

EFRA – Develop the first analytics-enabled, secure-by-design, green data space for AI-enabled food risk prevention. Our mission is to support EU's global leadership in the digital-led industry transition from reaction to food risk prevention. **efraproject.eu**

EXA4MIND – Building a platform for extreme data that enables advanced data analytics on supercomputers and automated data management with support for integration by design with EOSC and European data spaces. **exa4mind.eu**

DataNexus: How?

DATA NEXUS

Graph-Massivizer: Extreme and Sustainable Graph Processing for Urgent Societal Challenges in Europe.

Use Cases: Massive graph representation of extreme data in:

- Data Center Digital Twin for Sustainable Exascale Computing
- Green AI for Sustainable Automotive Industry
- Global Foresight for Environment Protection
- Green and Sustainable Finance

NEARDATA: Extreme Near-Data Processing Platform.

Use Cases: A near-data platform for consumption, mining and processing of data in:

- Genomics
- Metabolics
- Surgery

EXA4MIND: EX4MIND platform for extreme data enables advanced data analytics on supercomputers and automated data management with support for integration by design with European data ecosystems.

Use Cases:

- Systematic improvement of molecular simulations and accuracy
- Massively automated annotation and evaluation of automotive camera recordings
- Smart farming/viticulture with sensors and satellite imagery
- Secure data mining in health data

EXTRACT: A distributed data-mining software platform for extreme data across the compute continuum.

Use Cases:

- Personalized evacuation route system (mobile app)
- Transient astrophysics for analyzing space weather

SYCLOPS: Scaling extreme analytics with cross-architecture acceleration based on open standards, and advancing AI/data mining by democratizing its acceleration through open standards.

EFRA: Extreme Food Risk Analytics.

Use Cases:

- Risk predictions for poultry pathogens
- Enhanced predictive capabilities for pest alarms
- Informing regulatory decisions with food risk intelligence

EMERALDS: Extreme-scale Urban Mobility Data Analytics as a Service. The MAaaS toolset will stand out by moving analytics and sensitive data analytics to edge computing, enhancing response times and safeguarding data privacy.

Use Cases:

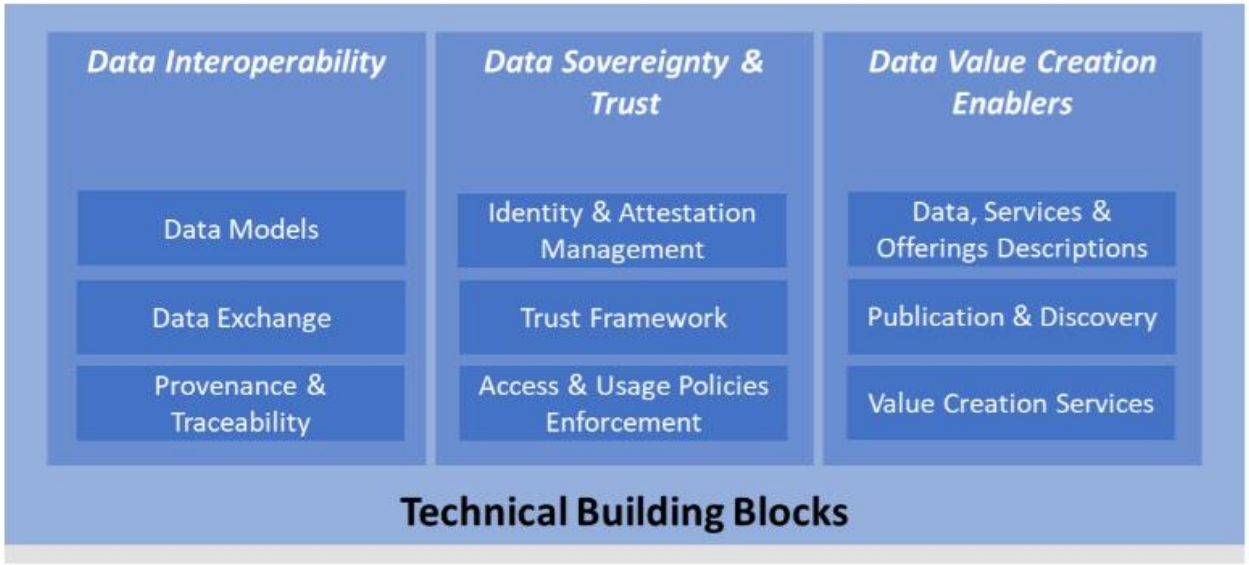
- Mobility analytics as a service toolset
- Risk-assessment, prediction and forecasting during events
- Multi-modal integrated traffic management
- Trip characteristics inference and traffic flow data analytics

Capture QRcode or follow this URL

horizonresultsbooster.eu

Synergies with data spaces and contributions to DSSC Blueprint

- Greater focus on the data value creation layer (analytics)
- Technologies suited to deal with extreme data
- Support to companies in processing data across the computing continuum (infrastructure layer)



Materials

- Press-release
- Flyer
- Materials generated for workshops (e.g. EBDVF)
- https://www.youtube.com/watch?v=CLBs7Si0MNo
- https://www.youtube.com/watch?v=7kTdhwvELB4&t=139s
- **White paper in the making**