



Federált tanulás és lehetőségei: COMPLIANCE BY DESIGN IN THE EU CLOUD

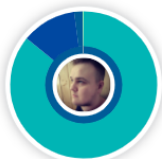
Tényi Ákos

E-Group ICT Software Group,
Head of Smart Data Business Line
akos.tenyi@egroup.hu

2023. November 08.





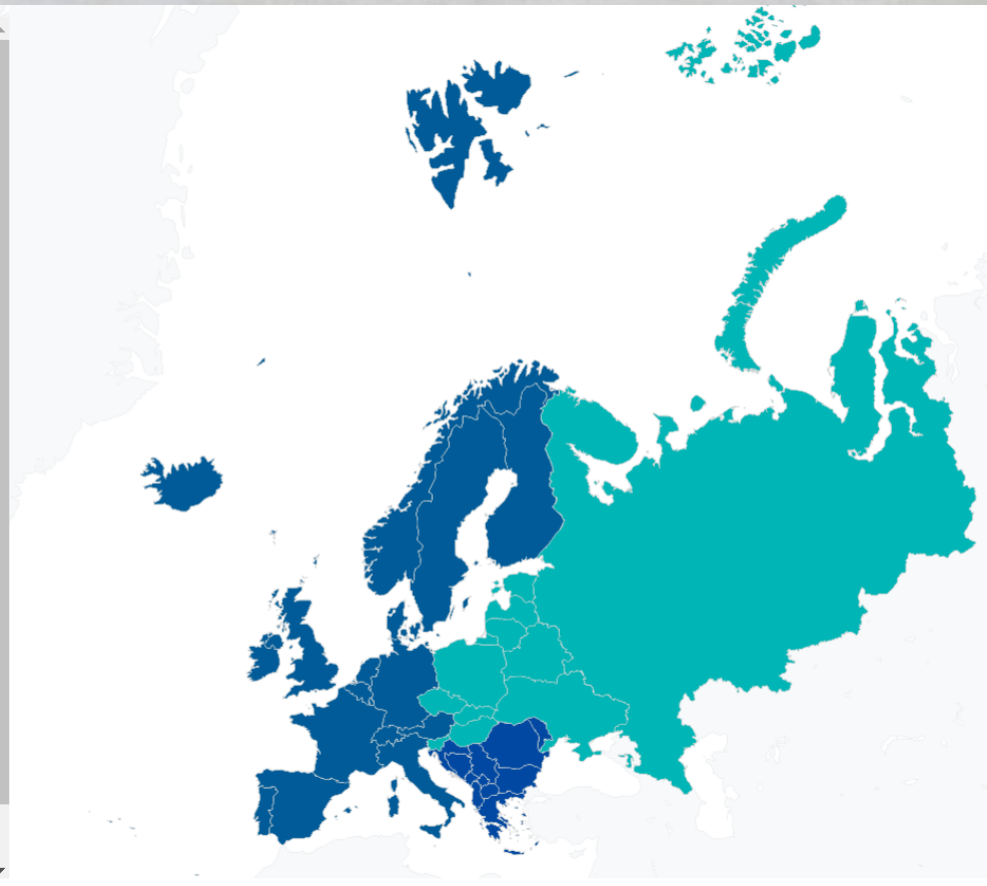


Gergő

100%

European		99.7%
● Eastern European		85.8% >
Podkarpackie Voivodeship, Poland		
Prešov Region, Slovakia		
+24 regions		
● Southern European		12.0% ▾
● Greek & Balkan		12.0% >
Alba County, Romania		
+9 regions		
● Broadly European		1.9% ▾
Trace Ancestry		0.2% ▾
Unassigned		0.1% ▾

Updated: June 14, 2021 ⓘ



1st NATIONAL FEDERATED BIOBANK IN EU

The **Semmelweis Biobank Network** keeps the data of approx. 100,000 individuals. This data asset is managed by diverse, fragmented, non-interoperable data collection systems.

We create **interoperability** between fragmented databases at the institutional level and to make it possible to jointly analyze them in a way that **guarantees the security of the individual's personal data**. We create **federated data assets at the institutional level** that preserves data protection guarantees, which supports **interactive cooperation between cross-border health data repositories**, the identification of new knowledge, a **better understanding of the pathomechanism of disease**, with the support of clinical research and decision-making. **Data sharing that ensures privacy will ensure optimal interoperability between data collection centers.**



Semmelweis University
leading universities of medicine and health
sciences



MAIN CHARACTERISTICS

TRANSPARENCY

INTEROPERABILITY

FEDERATED DATA EXCHANGE

F. Catalog

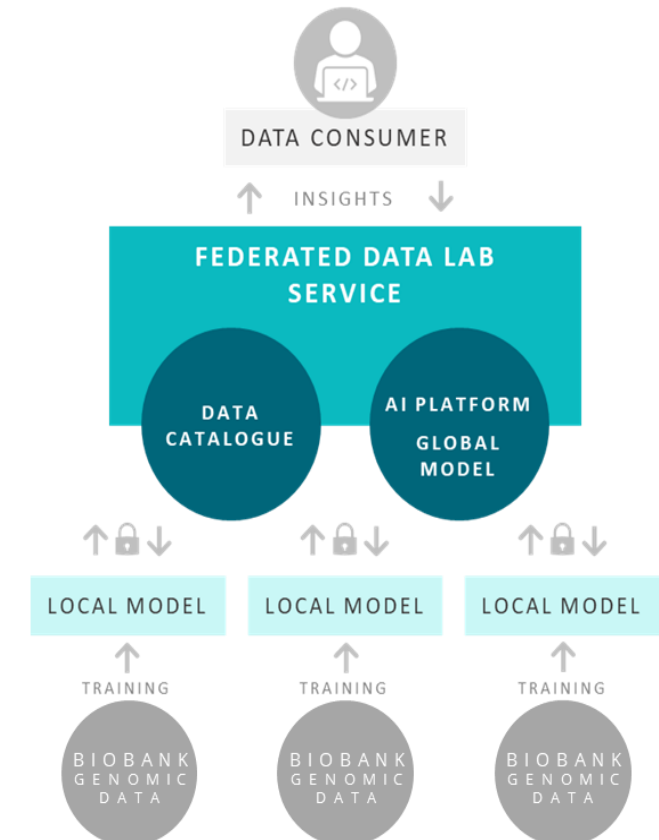
F. Cohort Discovery

F. Cohort Statistics

F. Cohort Study

HIGHLIGHTS

- **12 independent biobank** in the network
- FAIR data and data harmonization activities
- European data standards
- Secure and privacy assuring, distributed execution environment
- Analytical and virtual data exchange system
- Certificate enablement
- Testing and maintenance support



GLOBAL DATA INFRASTRUCTURE FOR AI & KNOWLEDGE

WHY DATA IS ESSENTIAL?

AI is top priority
strategy across the world



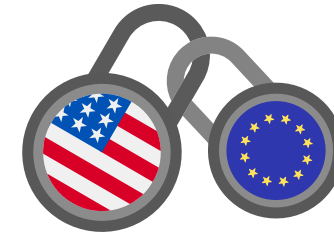
AI needs data

There is **no** modern and
productive industry without
data intensive operations



EU-US AI agreement:

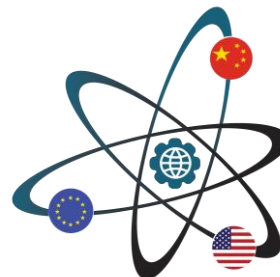
The U.S. data stays in the U.S. and
European data stays there, together
better models



Data is distributed across
countries.



Who owns the data + the data
extraction technology (AI) **has the**
control as well



GDPR, EU AI Act:

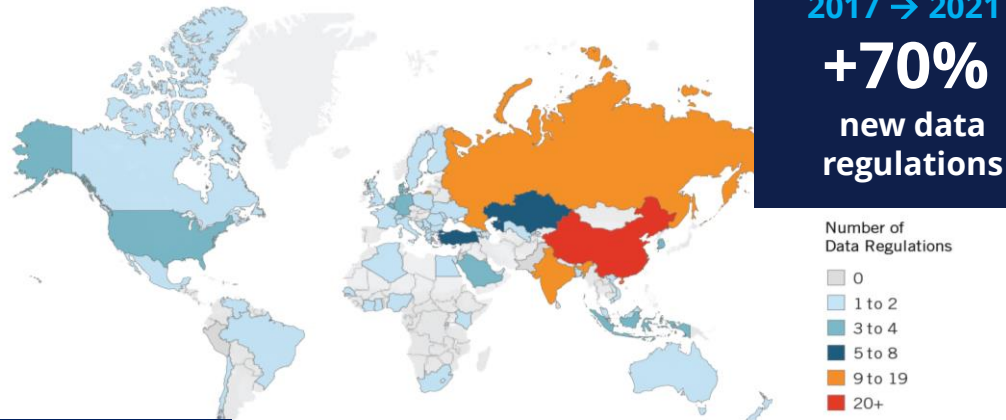
Training highly restricted, localised on EU
data

Emphasis on representative training,
testing data sets of AI models



LIMITS ON DATA USAGE AND ACCESS

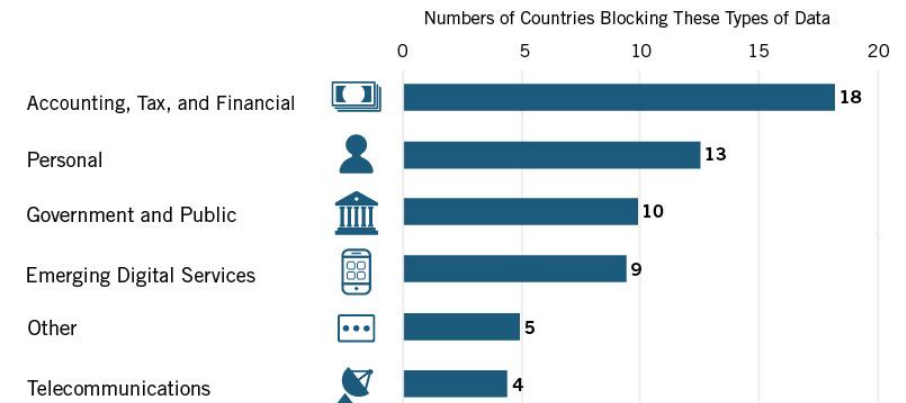
WHY TECHNOLOGICAL PROJECT IS ESSENTIAL?



**Data protection
laws**
EU: GDPR, AI Act

“Meta hit with record-breaking \$1.3 billion fine over Facebook data transfers to the US”

WHAT TYPES OF DATA ARE BLOCKED?



**ITIF analysis of formal laws or regulations publicly reported as of April 2017.*

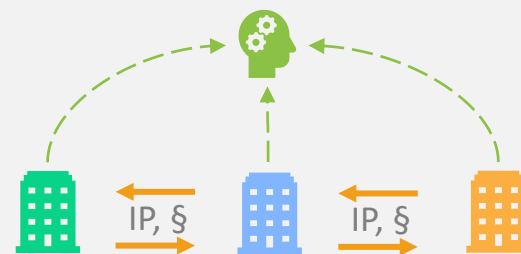
Learn more at itif.org/databarriers



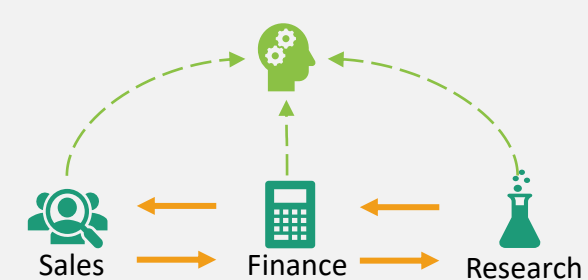
CROSS-BORDER



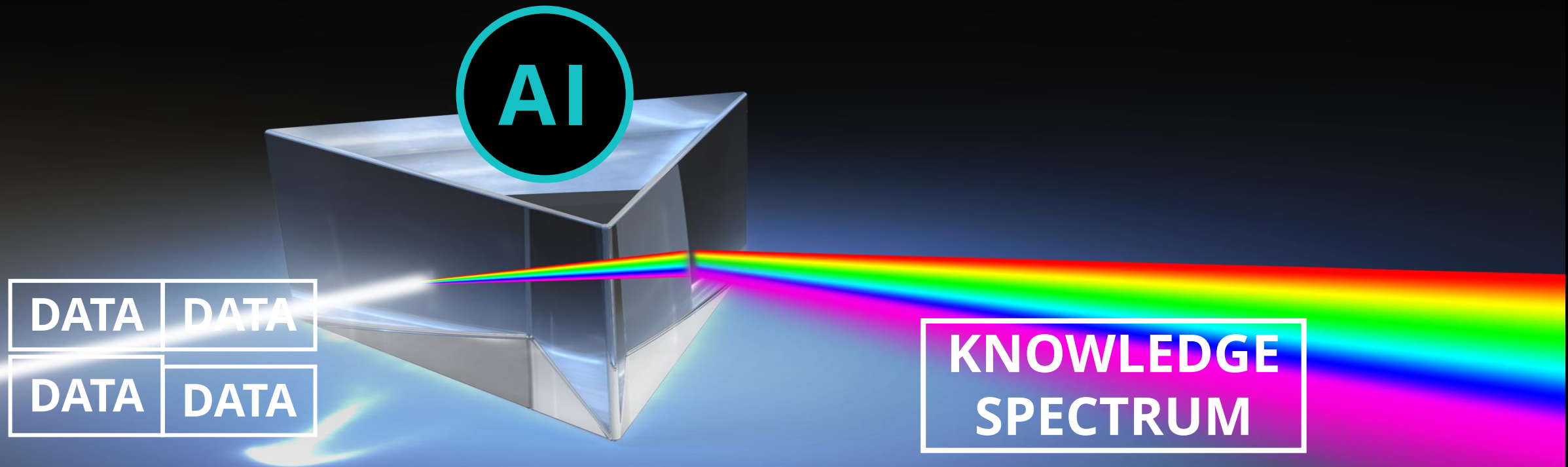
CROSS-ENTERPRISE



CROSS-DEPARTMENT



FROM DATA TO INSIGHT = KNOWLEDGE SPECTRUM



Giving out data , giving out the whole knowledge spectrum

Not "enough" data > knowledge inferior> Is „knowledge licencing" in narrow sense possible?

Monitoring and controlling of Insight-generation

A close-up photograph of a chimpanzee's face. The chimpanzee has dark black fur and a reddish-brown face. It is looking directly at the camera with a serious expression. Its right hand is raised to its forehead, with its fingers spread, as if shielding its eyes or in a gesture of deep thought. The background is a soft, out-of-focus green, suggesting a natural habitat.

WHAT TO DO?

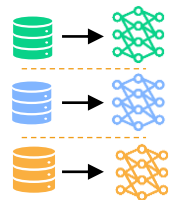
FEDERATED LEARNING= DIST.KNOWLEDGE GENERATION

BRINGS ANALYTICS TO THE LOCAL DATA AND
FEDERATES extracted KNOWLEDGE

- Brings the algorithm to the data
- No need to aggregate data
- Machine Learning models aggregated
- Private data kept at the place of origin
- **Privacy preserving** property of learning process ENSURED
- Secure and fast access to data

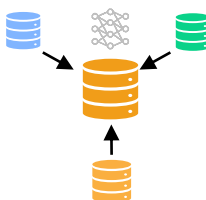
MACHINE LEARNING TECHNIQUES

A LOCAL LEARNING



Local data, local learning

B CENTRAL LEARNING

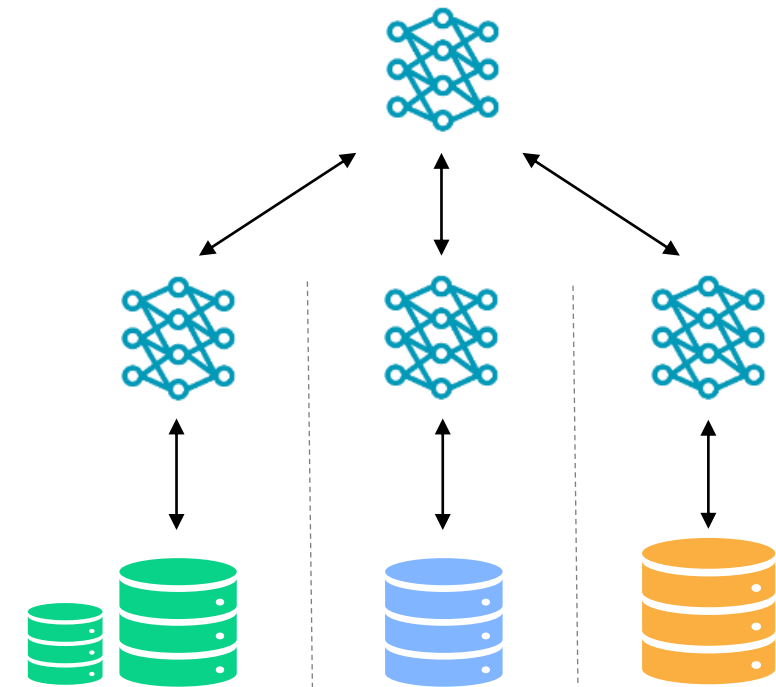


Centralized data, centralized learning

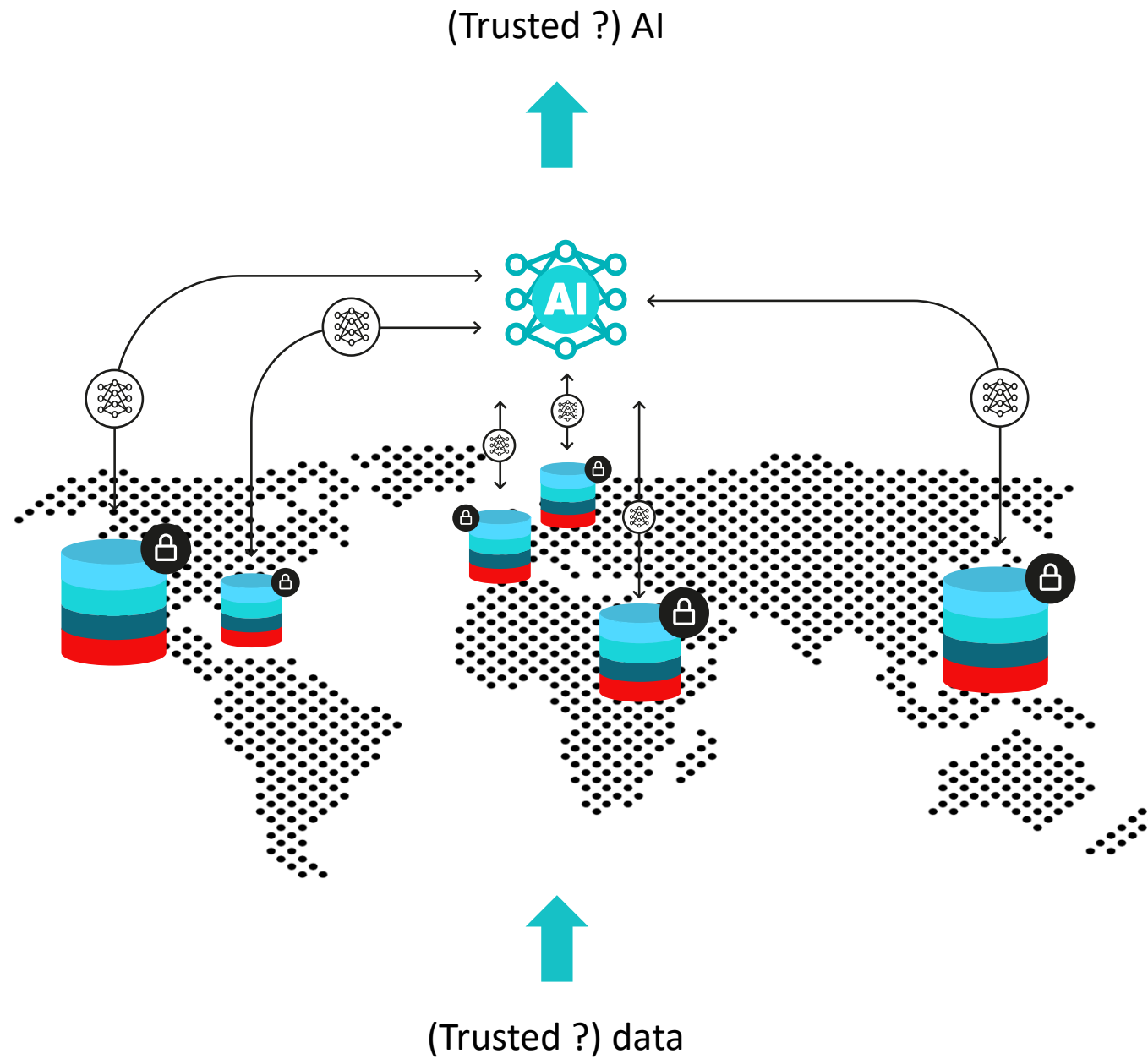
CONCEPT

Train machine learning algorithms across multiple decentralized servers (datasets) without sharing their data

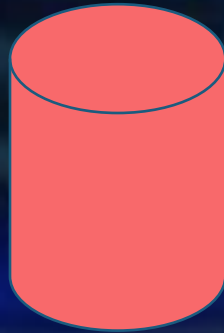
C FEDERATED LEARNING



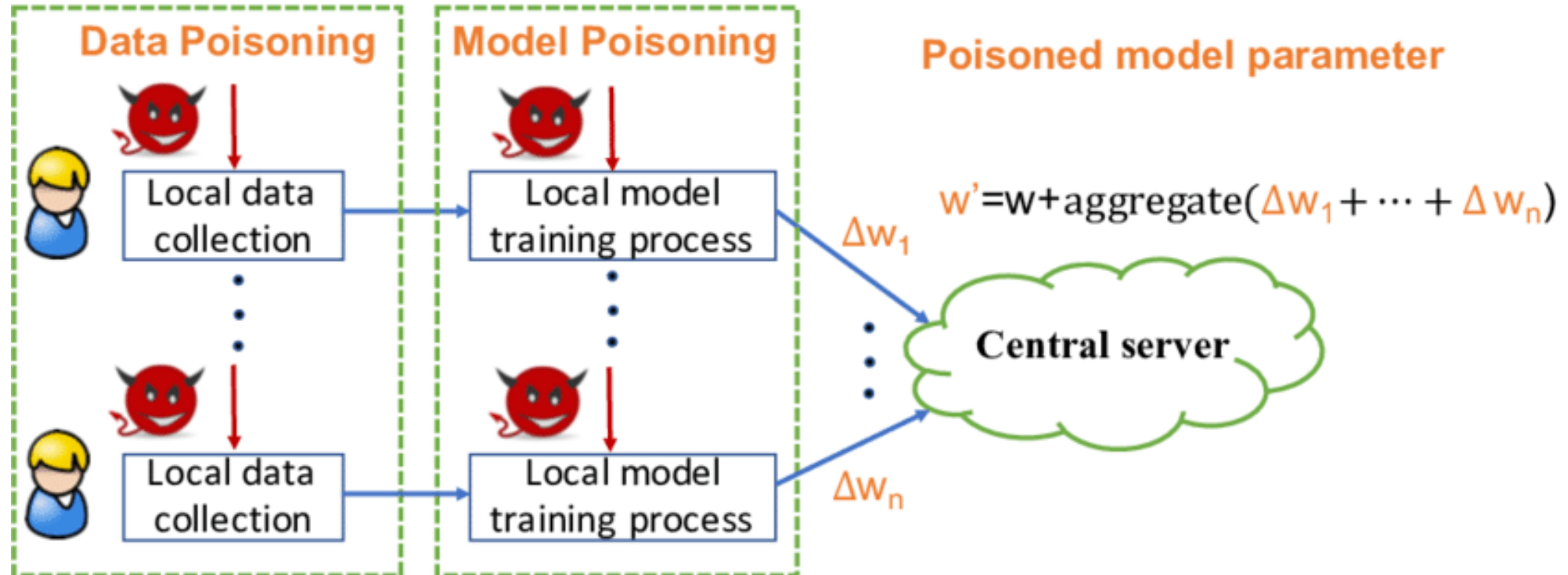
Local data
local learning + global learning



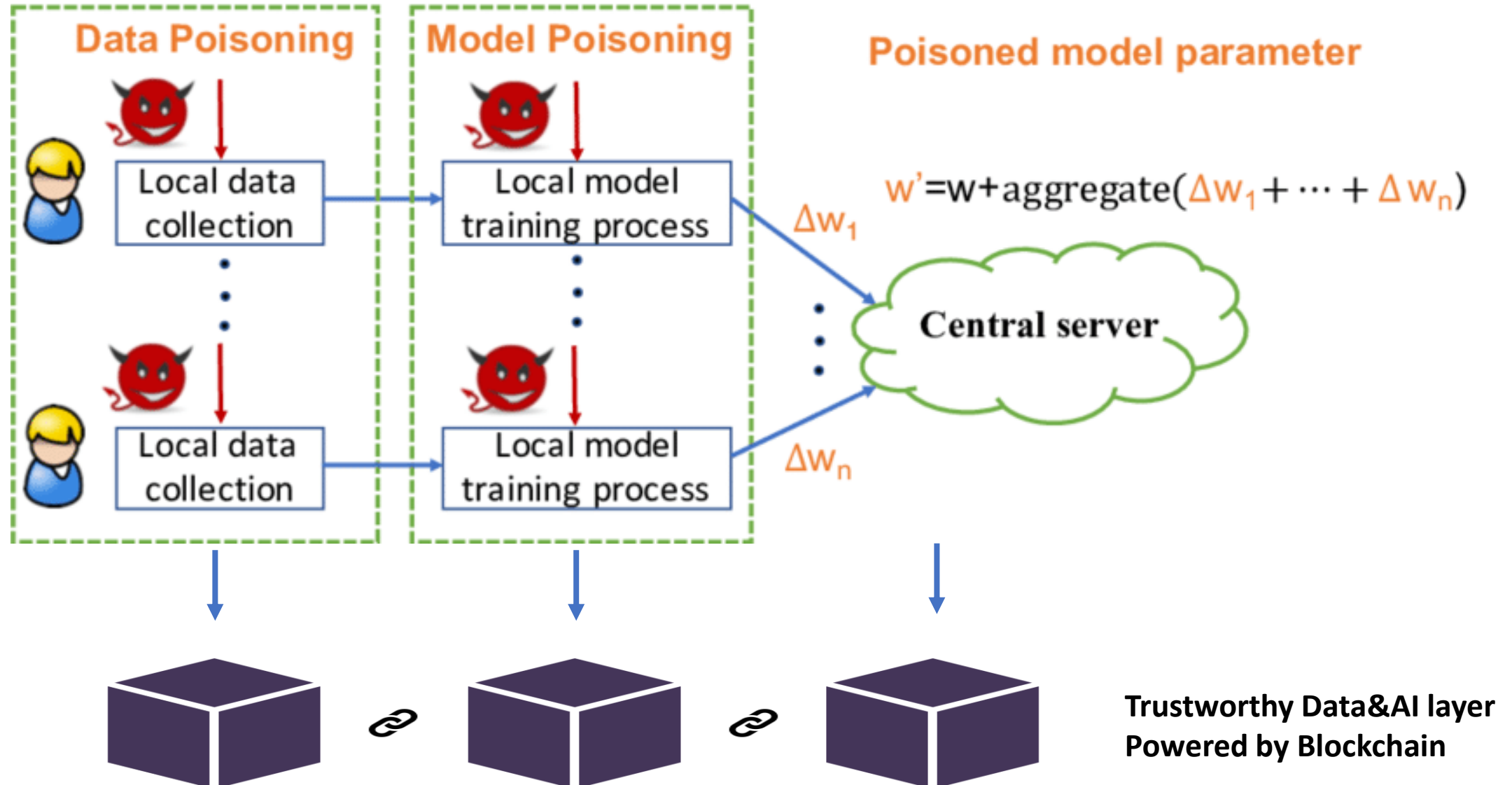
(Trusted ?) data



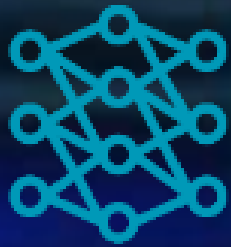
Trustworthy data?



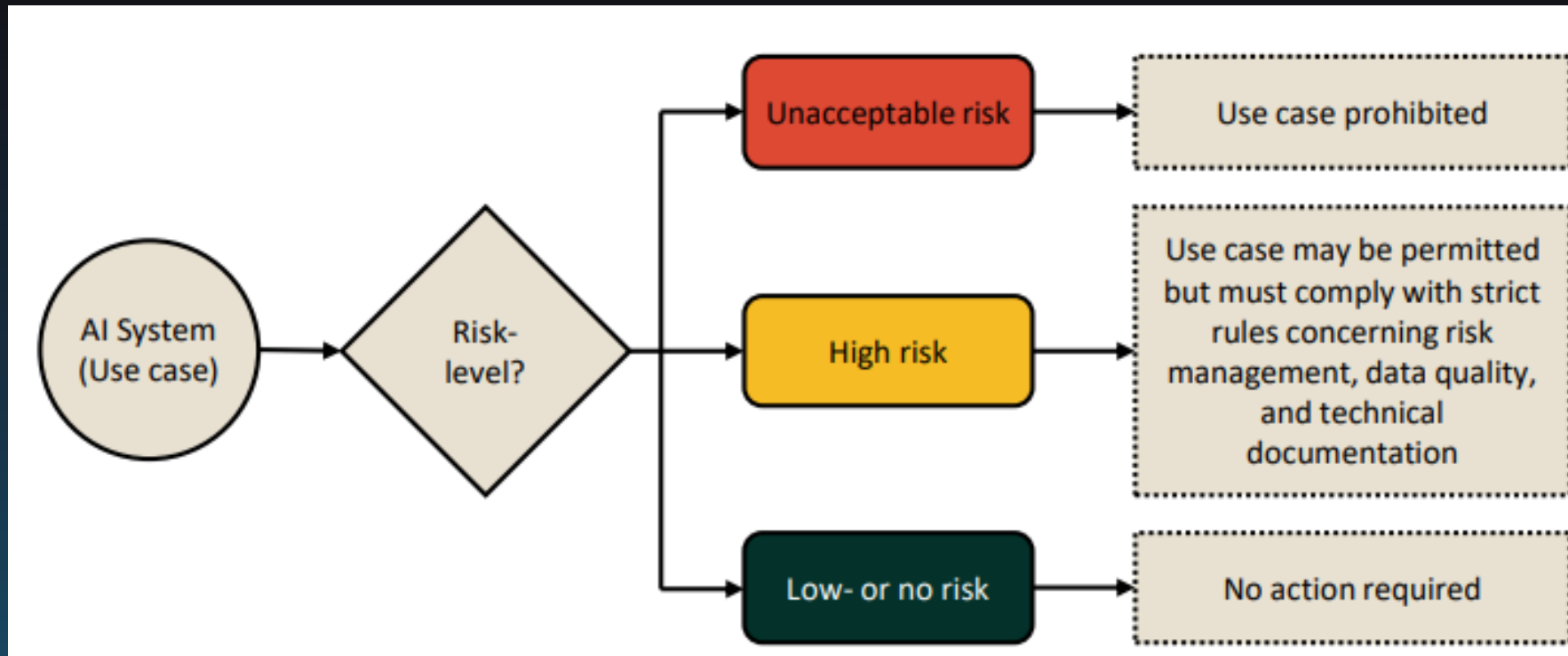
Need for trustworthy data and models



(Trusted ?) AI



AI ACT – Risk based classification



AI ACT – CONFORMITY NEEDS ADVANCED DATA STRATEGY

CONFORMITY ASSESSMENT

- **Data & data governance**

- **training, validation of model** based on
 - **relevant, representative data**
 - **w appropriate statistical properties** as regards to:
 - person or group
 - **geographical, behavioural or functional setting**
- bias monitoring, potentially based on personal data > **privacy preserving measures**

- **Model**

- **appropriate level of accuracy** of model (in light of intended purpose)
- robustness, fail-safe
- resilient to errors, unauthorised externals
- cybersecurity, content adversarial attacks
- **Intended purpose**
 - N x n (sectors x use cases) playbooks + situation specific
 - validation data needed



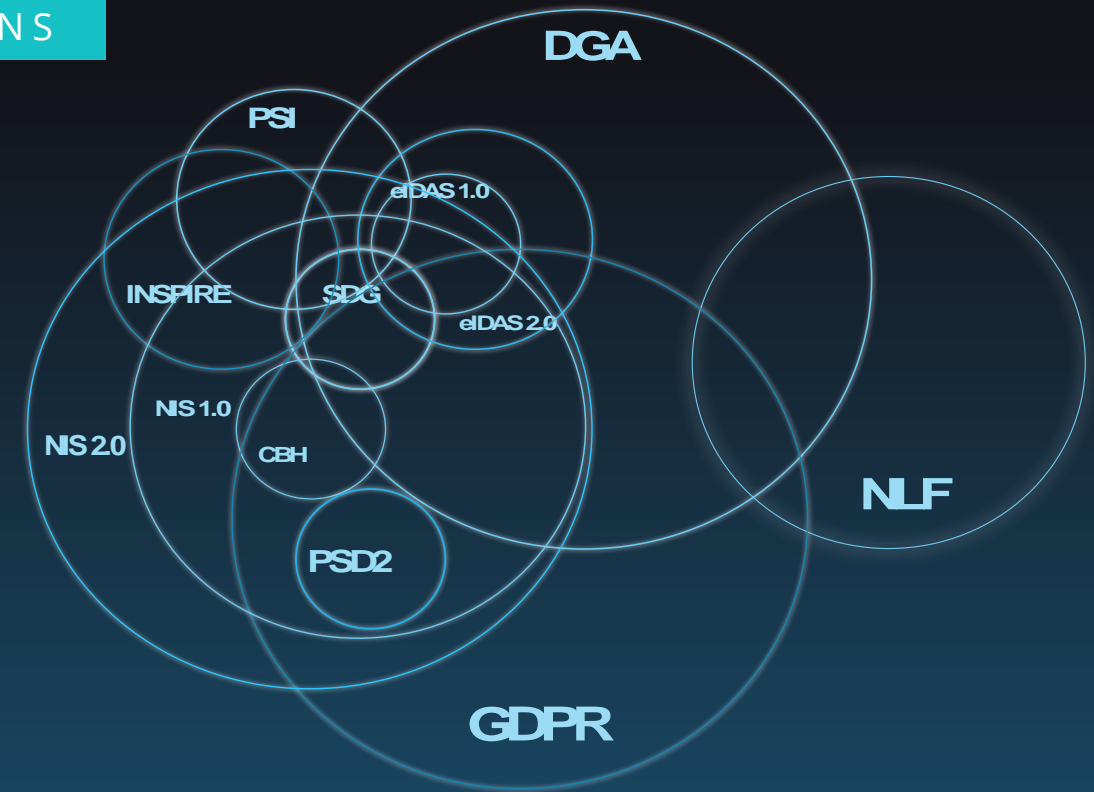
DATA STRATEGY IMPLEMENTATION REQUIRES
**ADVANCED DATA&AI LOGISTICS TECHNOLOGY
BACKING**

AI ACT - DATA REGULATION COMPLEXITY

COMPLEMENT/CONSISTENT WITH OTHER REGULATIONS

- General Data Protection Regulation (GDPR)
- Law Enforcement Directive (LED)
- Union law on non-discrimination
- Union competition law
- New Legislative Framework (NLF) (e.g. medical devices, machinery, toys)
- Financial services legislation (e.g. PSD2)
- E-Commerce Directive, Digital Services Act (DSA)
- closely linked to the Data Governance Act, the Open Data Directive and other initiatives under the EU strategy for data

re-use, sharing and pooling of data that are essential for the development of data-driven AI models of high quality



DATA REGULATION COMPLIANCE REQUIRES
CONTROLLED DATA PROCESSING ENVIRONMENTS

AI ACT - DATA REGULATION COMPLEXITY

COMPLEMENT/CONSISTENT WITH OTHER REGULATIONS

- General Data Protection Regulation (GDPR)
- Law Enforcement Directive
- Union law on non-discrimination
- Union competition law
- New Legislative Framework (NLF) (e.g. medical devices, machinery, toys)
- Financial services legislation
- E-Commerce Directive, Digital Services Act (DSA)
- closely linked to the Data Governance Act, the Open Data Directive and other initiatives under the EU strategy for data



re-use, sharing and pooling of data that are essential for the development of data-driven AI models of high quality



**DATA REGULATION COMPLIANCE REQUIRES
CONTROLLED DATA PROCESSING ENVIRONMENTS**

AI Regulatory Sandbox

Regulatory Sandbox shall provide a **controlled data processing environment** that facilitates the development, testing and validation of innovative AI systems for a limited time before their placement on the market

- contains personal data (lawfully collected for other purposes)
- aim to aid development and testing



REGULATORY SANDBOX "*BY DESIGN*"
REQUIRES **PRIVACY-FIRST DATA
PROCESSING TECH ENVIRONMENT**



E-GROUP
SOFTWARE & BEYOND

BLUEPRINT OF A REGULATORY SANDBOX

- **Req 1.** contains **highly relevant, private data** that needed for innovative AI development
- **Req 2. monitoring of risks** during development
- **Req 3.** functionally separate, isolated, **protected data processing environment**
- **Req 4.** only authorised person has access
- **Req 5. no personal data to be transmitted**, transferred or otherwise accessed by other participants
- **Req 6.** processing fully **preserves data subject privacy**
- **Req 7.** access to data can be fully revoked (time period, other reasons)

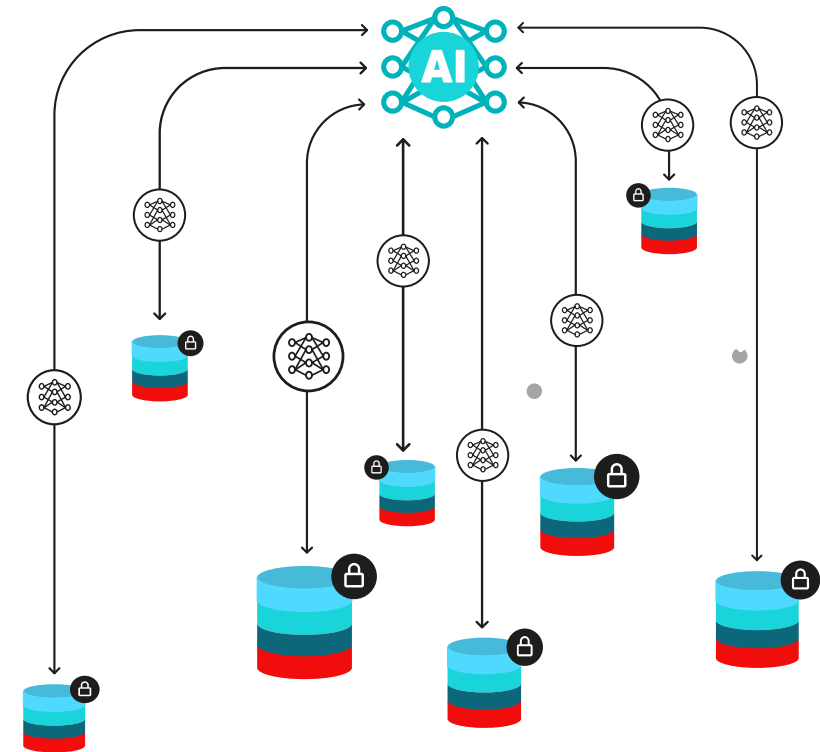
CREATE FEDERATED AI TECHNOLOGY & SERVICE ON DISTRIBUTED DATA NATIONAL, PAN-EU LEVEL

COMPETENT NATIONAL AUTHORITIES: for conformity assessment -> **Regulatory Sandbox!**

AI COMPANIES (SME & BIG Industry too): Development, testing, validation -> **AI compliance!**

AI COMPLIANCE SANDBOX

- **Trusted technology platform** for trustworthy data and AI development, validation EU level
- **Data stays, AI travels**
- Compliance with data protection laws and standardized, streamlined automatic processes
 - **Privacy by Tech Design**
 - Zero-risk “data processing” (data stays)
 - Standardized and streamlined data processing **security assessment**
 - Data security and **personal data rights guarantees embedded**
- Ability to support **National, Pan EU AI ecosystem**
- **Reduces complexity** and **enables new markets to thrive**
- Supports data enabled industry B2B AI/MI collaboration without **risking data sharing issues**



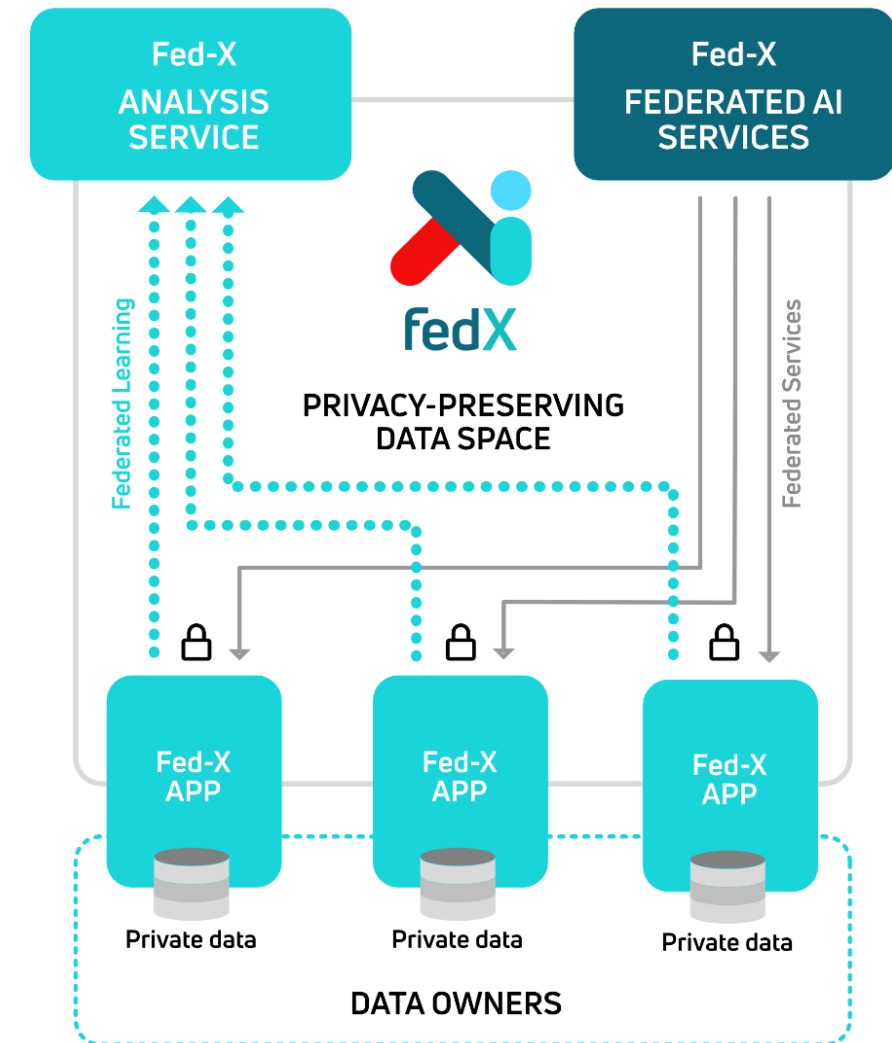
FEDX – FEDERATED AI LEARNING PLATFORM




PRIVACY FIRST, CONTROLLED, FEDERATED **AI SANDBOX FOR CONFORM AI TRAINING, TESTING AND VALIDATION.**

HIGHLIGHTS

- **Federated Client Database:** functionally separate, isolated, protected data processing environment (Req 1., Req 3. see earlier slide AI Act requirement)
- **Trusted Federated Machine Learning Engine:** no personal data transmitted, transferred during model development or validation (Req 5.)
- **Privacy-preserving AI:** processing fully preserves data subject and data provider privacy (Req 6.)
- **Training & Resource Monitoring, Logging:** monitoring of risks during development (Req 2.)
- **User Management:** only authorized person has access, temporary access to data processing can be given (Req 4., Req 7.)
- **FedX SF®:** federated technology without adaptation barrier: zero loss of model accuracy, even on non-IID data, highly energy efficient training



Tech edge

Technology	 FedX SF® Federated Learning	Federated Learning	Centralized Learning	Homomorphic Encryption	Secure Multi-Party Computation
Product using technology	FedX	Owkin Connect	Amazon SageMaker	ZeroReveal (Enveil)	Sharemind
Data Privacy	High	High	Low	High	High
Computational Efficiency	High	Low	High	Low	Low
Communication Efficiency	High	Low	High	High	Low
Performance on Real-World Data	High	Medium	High	High	High
Machine Learning/AI Capabilities	High	High	High	Medium	Medium
Data Analytics Capabilities	High	Low	High	Low	Medium

Our TECH EDGE

FEDERATED VS. CENTRAL TRAINING ACCURACY

Accuracy Loss (compared to central)	Benchmark Data	Real World Data
State of the Art Federated Learning	3%	9%
FedX SF technology®	0%	0%



0% LOSS ON ACCURACY

- eliminates technology adaptation barriers
- QoS comparable to non-federated approach
- Works seamlessly with non-IID (independent and identically distributed, ~ Real world data)

FEDERATED COMPUTATION COST

TOTAL COMPUTATIONAL FACTOR	Benchmark Data	Real World Data
State of the Art Federated Learning	18	54
FedX SF technology®	4	5



90% ENERGY SAVINGS vs Federated learning solutions

- Low cost OPEX
- Competitive pricing
- Green Incentives

EUROPEAN FEDERATED CLOUD DATA INFRASTRUCTURE

THE FEDEU.AI PROJECT CANDIDATE

- **Privacy preserving** technology platform for ML training
- **Scalable, cloud-native federated learning system**
- Advanced Federated Learning features, data catalogue and libraries
- Advanced **security & privacy protocols**
- Reducing cost and energy needs of federated training
- Technology platform for trustworthy data and models across
- **Reduces complexity and enables new market**



IPCEI-CIS

IMPORTANT PROJECTS OF COMMON EUROPEAN INTEREST

IPCEI ON NEXT GENERATION CLOUD INFRASTRUCTURE AND SERVICES

12 EU Member States join forces to create a common cloud and edge infrastructure and its associated smart services for the future

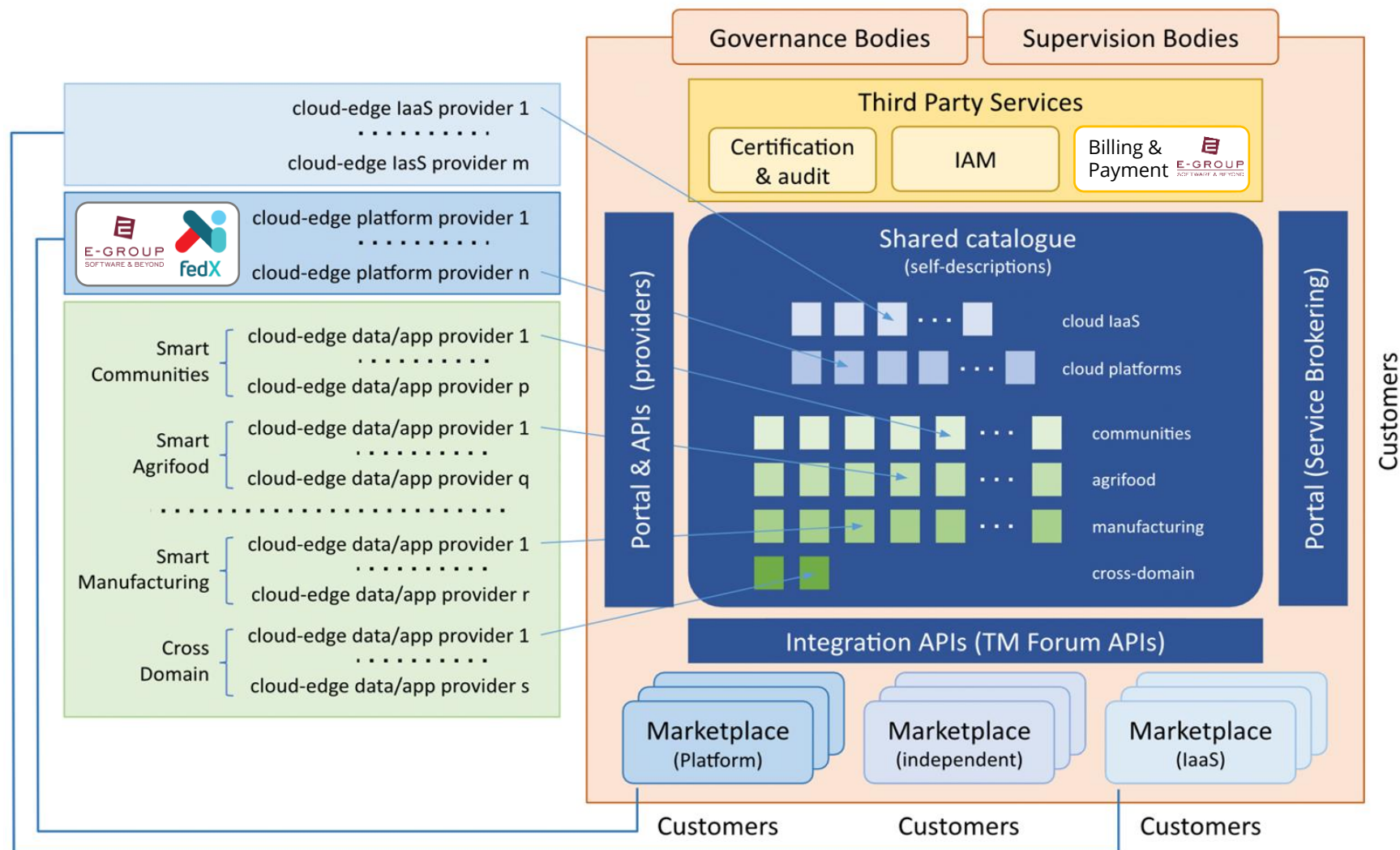




DOME EU Project (E-Group FedX POC)

A DISTRIBUTED OPEN MARKETPLACE FOR EUROPE CLOUD AND EDGE SERVICES (2023-2025, Project)

CLOUD-EDGE
PLATFORM PROVIDER



E-Group is a science and idea-driven software technology and digital knowledge manufacturing corporation taking great pride in its deep academic roots.

We strive to build a better future leveraging knowledge and wisdom derived from data.



E-GROUP IS

TRANSFORMATION SOLUTION
MULTIGENERATION INVENTIVENESS
ENERGY TECHNOLOGY CREATIVITY
RESULTS FAMILY
INFORMATION SECURITY QUALITY PARTNER
TALENTED ACADEMIC TEAM EXECUTION
WISDOM EXCELLENCE

TRUST | XCHANGE | DATA

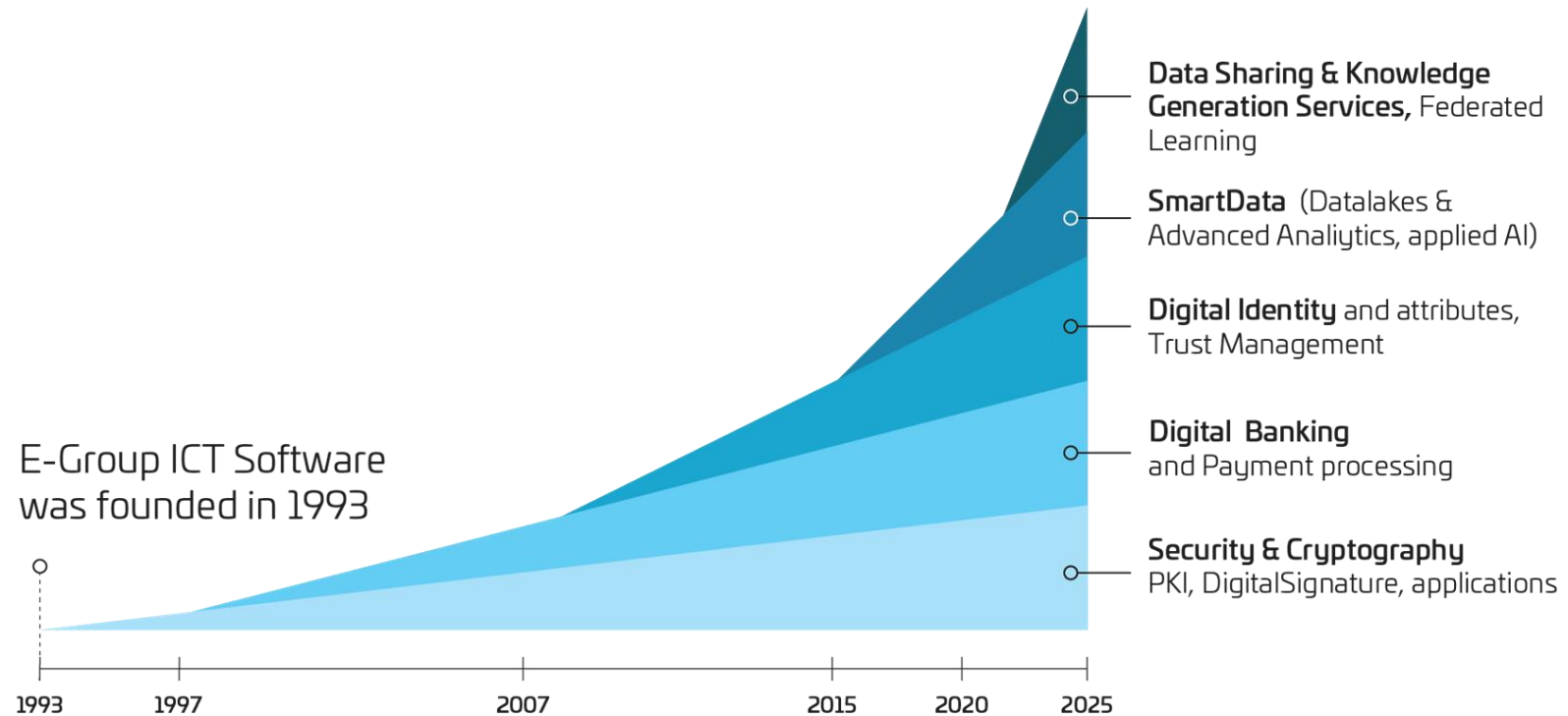
Our successful local and international projects in the FinTech, GovTech, HealthTech and EnergyTech segments prove that with our dedication and expertise, we create outstanding value for our clients and partners.

ALMOST 30 YEARS OF COMPETENCY GROWTH

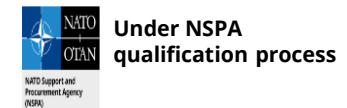
#TRUST

#XCHANGE

#DATA



ALLIANCES AND CERTIFICATES

**ESA Registered**

E-GROUP PARTNERS WORLD MAP

PARTNERSHIPS ARE ESSENTIAL
TO E-GROUP'S SUCCESS

We appreciate our cooperation with all
of our academic, technology and
innovation partners.



#KEY INDUSTRY PARTNERS



#EU NETWORK:: DATA AI AND INNOVATION ECOSYSTEM PARTNERS



R&D ACADEMIC PARTNERS





fedX



Antal KUTHY
E-Group CEO & Founder
antal.kuthy@egroup.hu

Ákos TÉNYI PhD
E-Group, Head of SmartData
akos.tenyi@egroup.hu