BEYOND AI ACT:
COMPLIANCE BY DESIGN IN THE EU CLOUD

Antal KUTHY
E-Group ICT Software Group, CEO & Founder
antal.kuthy*egroup.hu (replace * with @)

June 28. 2023

E-GROUP
SOFTWARE & BEYOND

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.

SOLUTION
TRANSFORMATION
INVENTIVENESS
MULTIGENERATION
TECHNOLOGY
CREATIVITY
ENERGY
E-GROUP IS
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

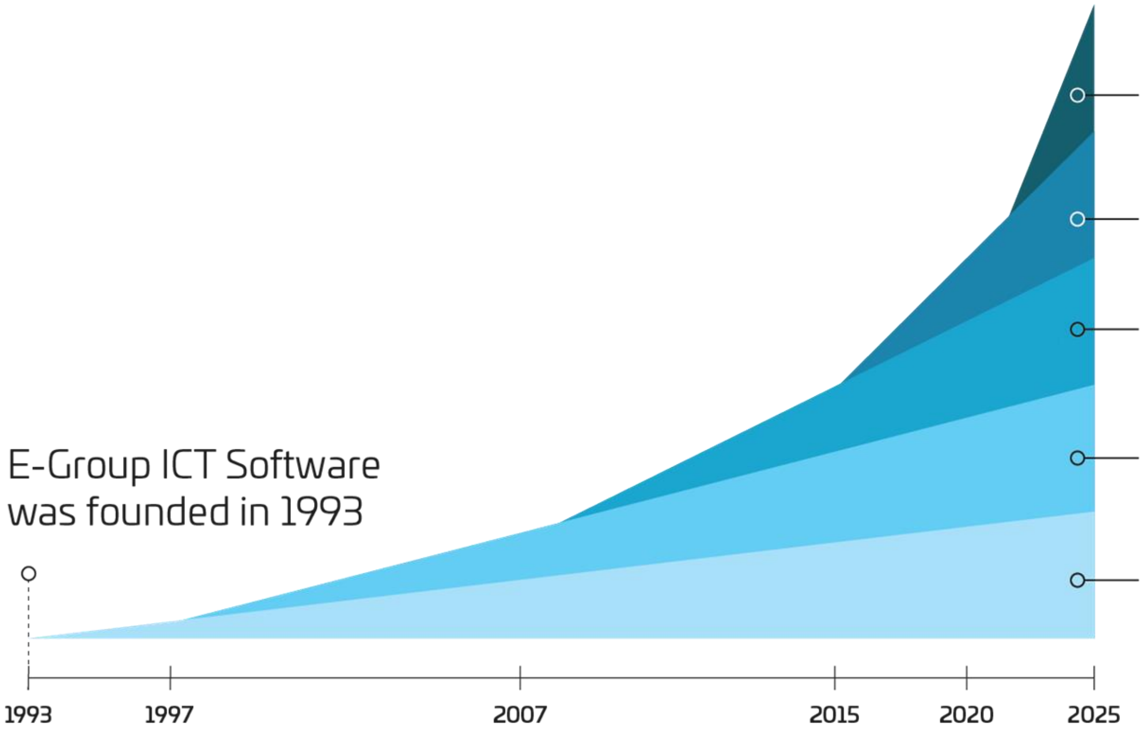
E-GROUP
SOFTWARE & BEYOND

#TRUST

#XCHANGE

#DATA

E-Group ICT Software was founded in 1993



Data Sharing & Knowledge Generation Services, Federated Learning

SmartData (Datalakes & Advanced Analiytics, applied AI)

Digital Identity and attributes, Trust Management

Digital Banking and Payment processing

Security & Cryptography PKI, DigitalSignature, applications

ALLIANCES AND CERTIFICATES

IPCEI CIS & Health

ESA Registered

AI artificial intelligence coalition

eit Digital

eit Health

ISO 9001:2015
ISO/IEC 27001:2013

NATO OTAN NATO Support and Procurement Agency (NSPA) Under NSPA qualification process

E-GROUP ICT SOFTWARE FEDX

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.

INDUSTRY CLIENTS/ CO-INNOV PARTNERS

THALES F-Secure Roche

Bittium AtoS SIA CA

otpbank Liber bank UnionPay 银联

#EU NETWORK:: DATA AI AND INNOVATION ECOSYSTEM we interact

BBMRI Biobanking and Biomolecular Resources Research Infrastructure

IPCEI CIS & Health

gaia-x

EHDEN EUROPEAN HEALTH DATA & EVIDENCE NETWORK

eit Digital

eit Health

R&D ACADEMIC NETWORK

KTH VETENSKAP OCH KONST

Stockholm University

TU berlin TECHNISCHE UNIVERSITÄT BERLIN

ELTE EÖTVÖS LORÁND TUDOMÁNYEGYETEM

MŰEGYETEM 1782

Semmelweis University

PÉCSI TUDOMÁNYEGYETEM UNIVERSITY OF PÉCS

University of Veterinary Medici Budapest

清华控股 TSINGHUA HOLDINGS

MISKOLCI EGYETEM UNIVERSITY OF MISKOLC

E-GROUP ICT SOFTWARE FEDX

GLOBAL DATA INFRASTRUCTURE FOR AI & KNOWLEDGE

E-GROUP
SOFTWARE & BEYOND

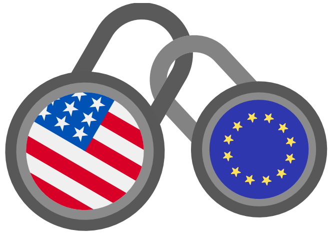
WHY DATA IS ESSENTIAL?

AI is top priority strategy across the world

AI needs data

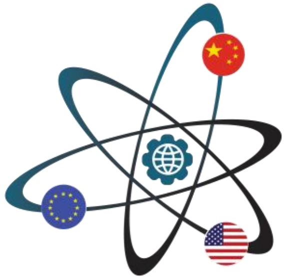
There is **no** modern and **productive industry without data intensive operations**

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



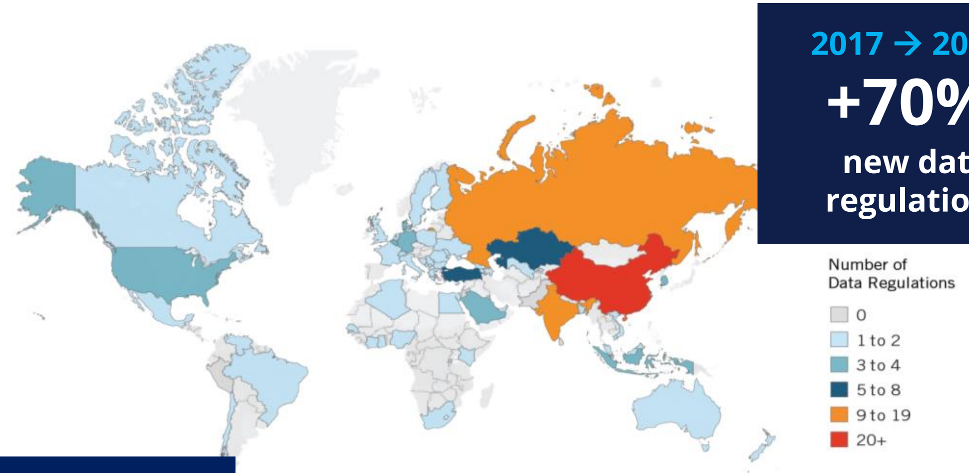
GDPR, EU AI Act:
Training highly restricted, localised on EU data
Emphasis on representative training, testing data sets of AI models

LIMITS ON DATA USAGE AND ACCESS

E-GROUP
SOFTWARE & BEYOND

WHY TECHNOLOGICAL PROJECT IS ESSENTIAL?

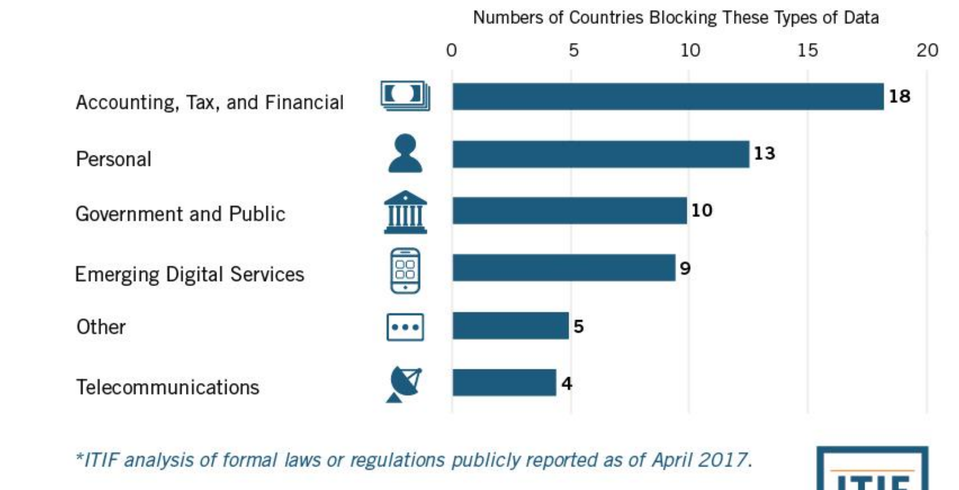
2017 → 2021
+70%
new data regulations



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?

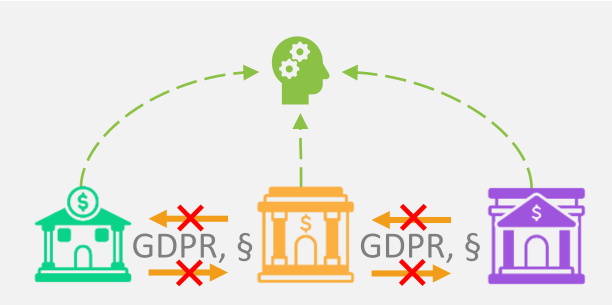


ITIF analysis of formal laws or regulations publicly reported as of April 2017.

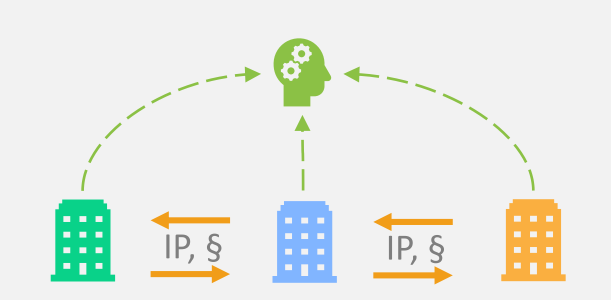
Learn more at itif.org/databarriers

ITIF INFORMATION TECHNOLOGY & INNOVATION FOUNDATION

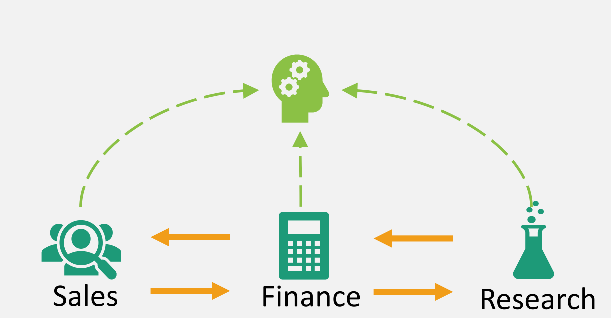
CROSS-BORDER



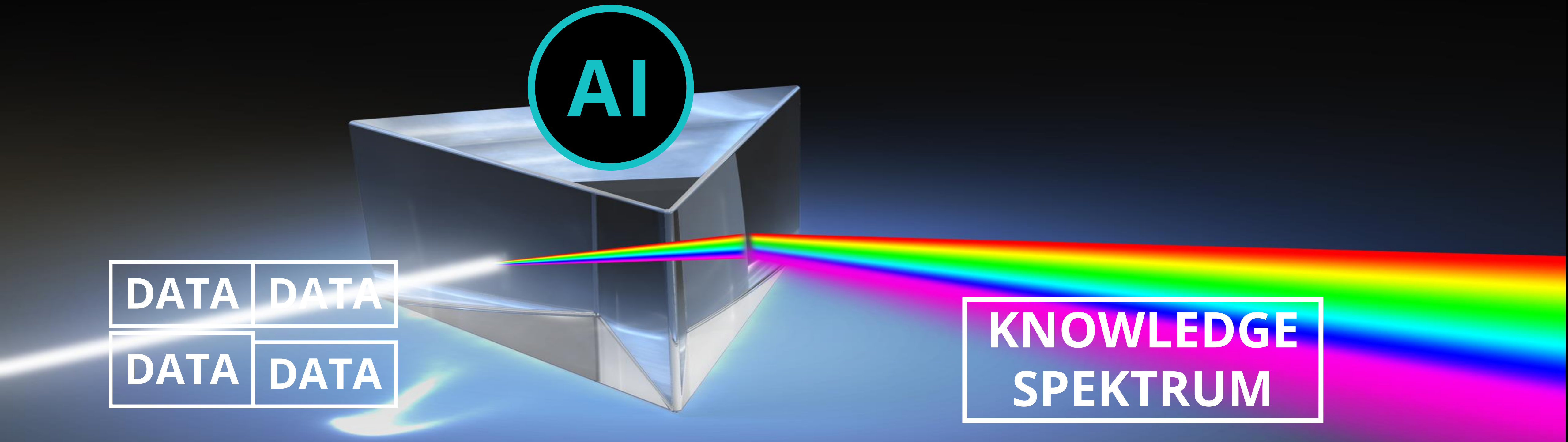
CROSS-ENTERPRISE



CROSS-DEPARTMENT



FROM DATA TO INSIGHT = KNOWLEDGE SPECTRUM



Gving out data , giving out the whole knowledge spectrum

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

Monitoring and controlling of Insight-generation

E-GROUP
SOFTWARE & BEYOND

ONE ADDITIONAL FACTOR: 'CHATGPT' PHENOMENON AND ITS CONSEQUENCES

(AND SIMILAR GENERATIVE AI SERVICES OFFERED GLOBALLY)

WHAT ARE THE KEY NEW FACTOR FROM REGULATORY POINT OF VIEW?

PROBLEM 1

1.**A few controls a global „knowledge distillation mechanism"**

2. **many contributes to the knowledge** by contributing knowledge via data

3. **not symmetric economic value re-distribution**

4. exponentiality in the gap

PROBLEM 2

Everyone is injecting knowledge into a centralised knowledge, more ridiculously pays for it – if does not want to remain competitive – forced adaptation

Exponential central knowledge repository growth > Centrally enabled intelligence becoming much higher than local knowledge driven intelligence > Innovative format of knowledge colonisation without physical brain drain

Scale is creating divide, industry **competitiveness and values** **industry diversity**

ONE ADDITIONAL FACTOR: 'CHATGPT' PHENOMENON AND ITS CONSEQUENCES

(AND SIMILAR GENERATIVE AI SERVICES OFFERED GLOBALLY)

WHAT ARE THE KEY NEW FACTOR FROM REGULATORY POINT OF VIEW?

PROBLEM 3

Even if an AI service promises not to store the data/information injected/input (we put in) but the trained model (e.g. LLM) reflects all the local assembled local knowledge in a generalised central knowledge even if I want withdraw my consent it is **impossible** to separate knowledge micro-spectrum in the trained model generated from my injected data/knowledge. Knowledge is harvested and locked.

I CAN NOT WITDRAW my contributed knowledge from the global knowledge pool, even if I want.

THERE IS NO UNDO mechanism.

Individuals and local level industry players loose their ability to control IP/knowledge superiority etc. by using the global stack, regulator is unable to study, regulate, since lacking knowledge, and sandboxed regulatory playground.

E-GROUP
SOFTWARE & BEYOND

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

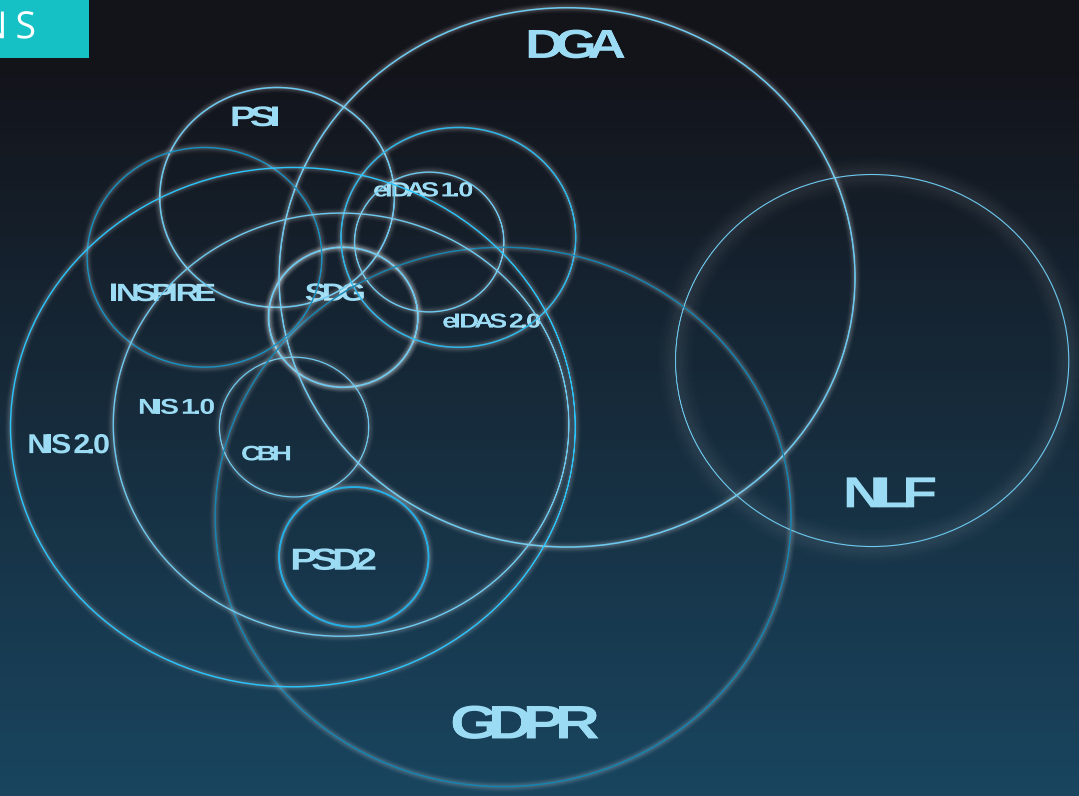
- Model
  - **appropriate level of accuracy** of model (in light of intended purpose)
  - robustness, fail-safe
  - resilient to errors, unauthorised externals
  - cybersecurity, content adversial attacks
- Intended purpose
  - N x n (sectors x use cases) playbooks + situation specific
  - validation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



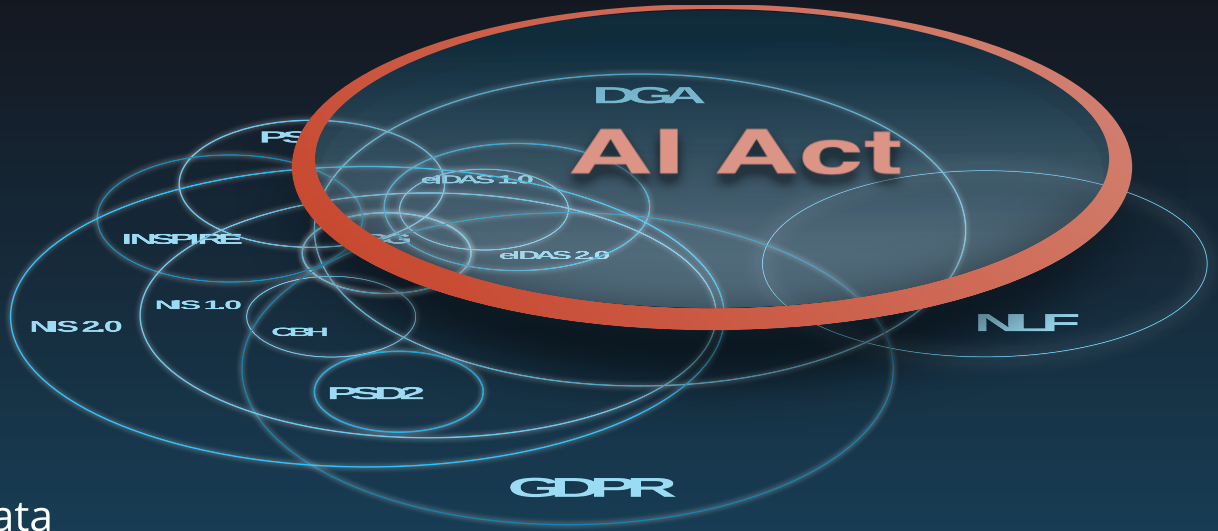
re-use, sharing and pooling of data that are essential for the development of data-driven AI models of high quality

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)

WHAT TO DO?

E-GROUP SOFTWARE & BEYOND

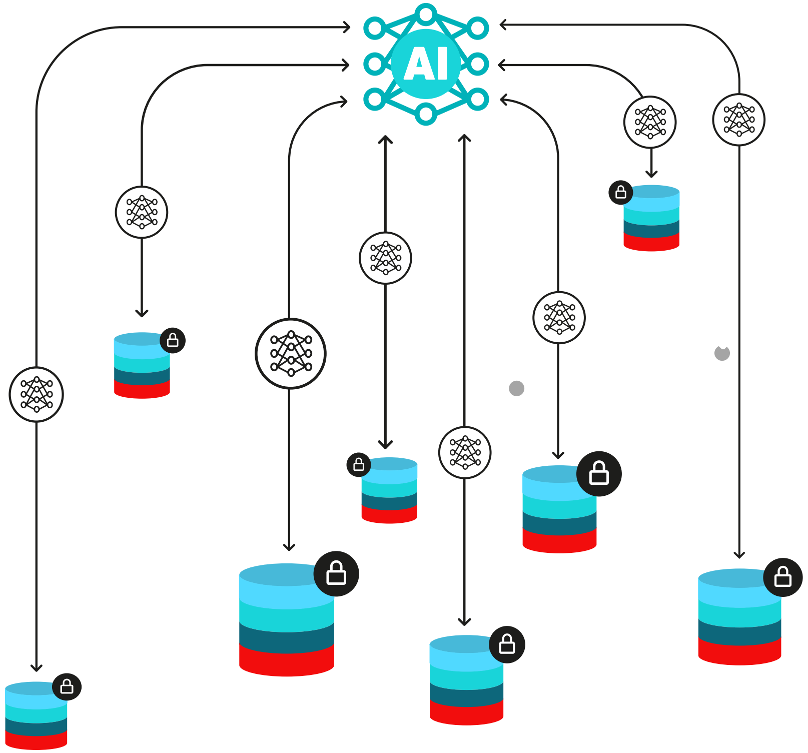
CREATE FEDERATED AI TECHNOLOGY & SERVICE ON DISTRIBUTED DATA NATIONAL, PAN-EU LEVEL

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

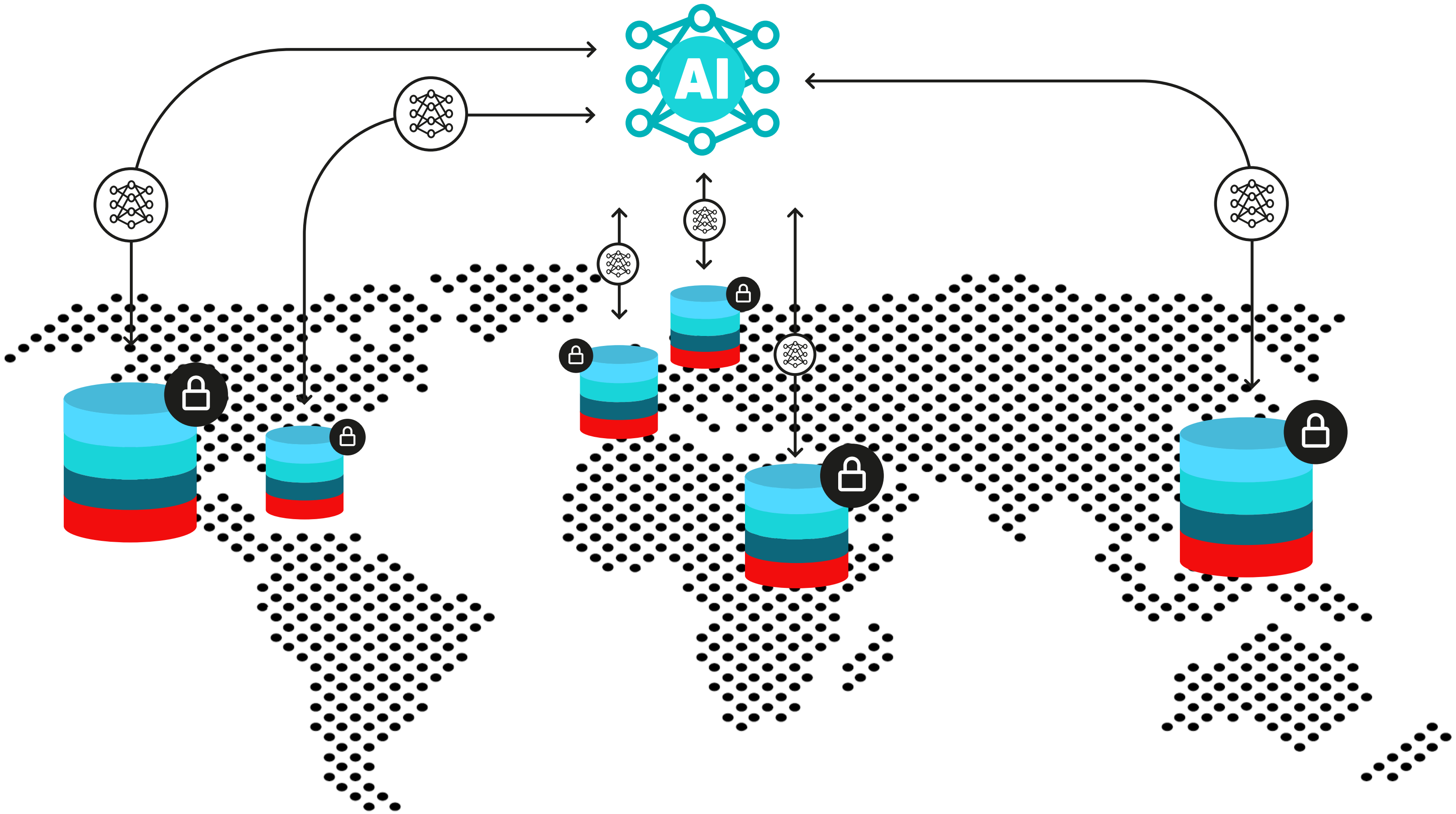
AI COMPANIES (SME & BIGIndistry too): Development, testing, validation -> **AI compliance!**

AI COMPLIANCE SANDBOX

- **Trusted technology platform** for t**rustworthy 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 strive**
- Supports data enabled industry **B2B AI/MI collaboration** **without risking data sharing issues**



fedX



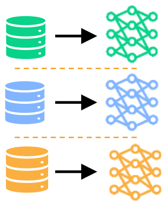
FEDERATED LEARNING= DIST.KNOWLEDGE GENERATION

BRINGS ANALYTICS TO THE LOCAL DATA AND FEDERATRES 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

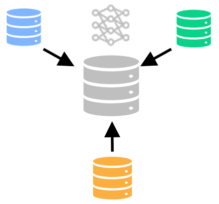
MACHINE LEARNING TECHNIQUES

A LOCAL LEARNING



Local data, local learning

B CENTRAL LEARNING

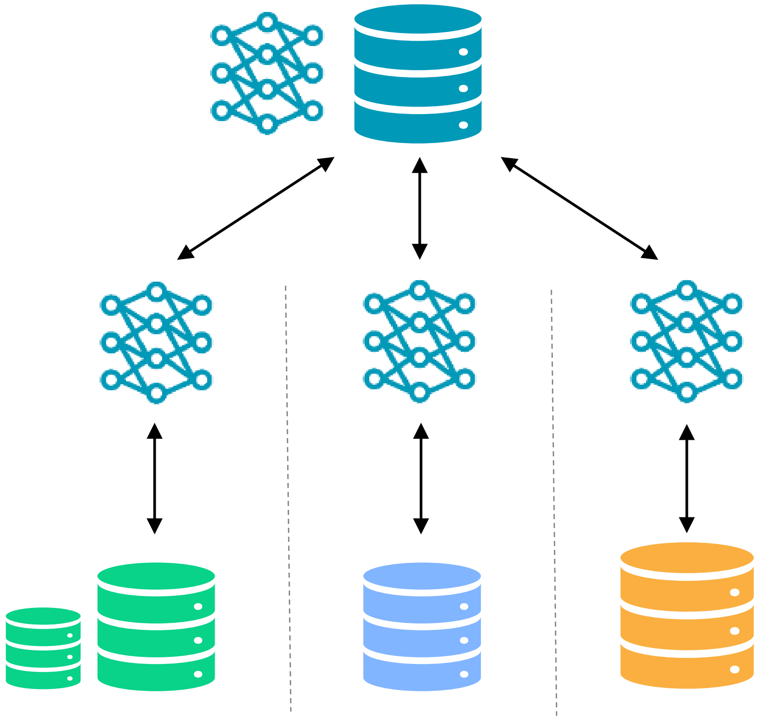


Centralized data, centralized learning

CONCEPT

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

C FEDERATED LEARNING



Local data
local learning + global learning

FedX – FEDERATED AI LEARNING PLATFORM **v1.0**

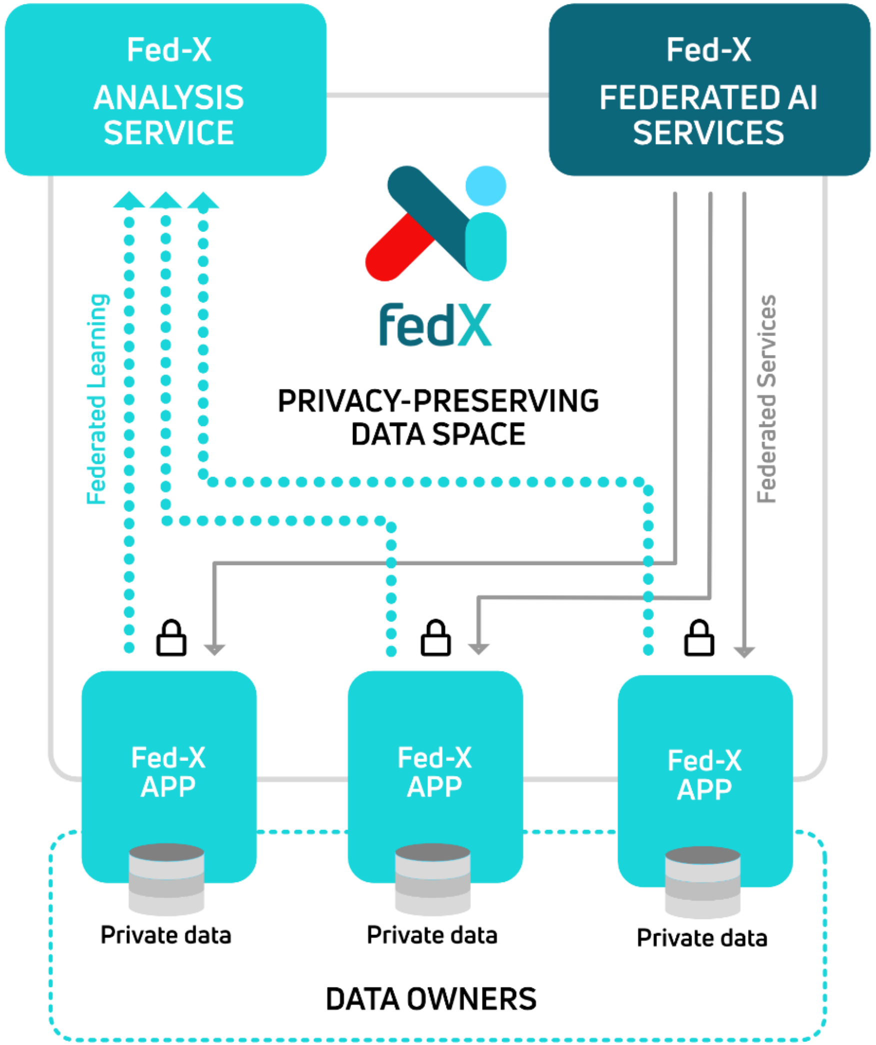
E-GROUP
SOFTWARE & BEYOND

fedX

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



Our FedX TECH EDGE

E-GROUP
SOFTWARE & BEYOND

FEDERATED VS. CENTRAL TRAINING ACCURACY

Accuracy Loss (compared to central)	Benchmark Data	Real World Data
State of the Art Federater 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 (idependent and identically distribute, ~ Real world data)

FEDERATED COMPUTATION COST

TOTAL COMPUTATIONAL FACTOR	Benchmark Data	Real World Data
State of the Art Federater Learning	18	54
FedX SF techology®	**4**	**5**

90% ENERGY SAVINGS

- ⇢ Low cost OPEX
- ⇢ Competitive pricing
- ⇢ Green Incentives

®FedX SF techNOLOGY is submitted to PCT/EPO patenting process

SUCCESS STORIES

EUROPEAN FEDRATED CLOUD DATA INFRASTRUCTURE

E-GROUP
SOFTWARE & BEYOND

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

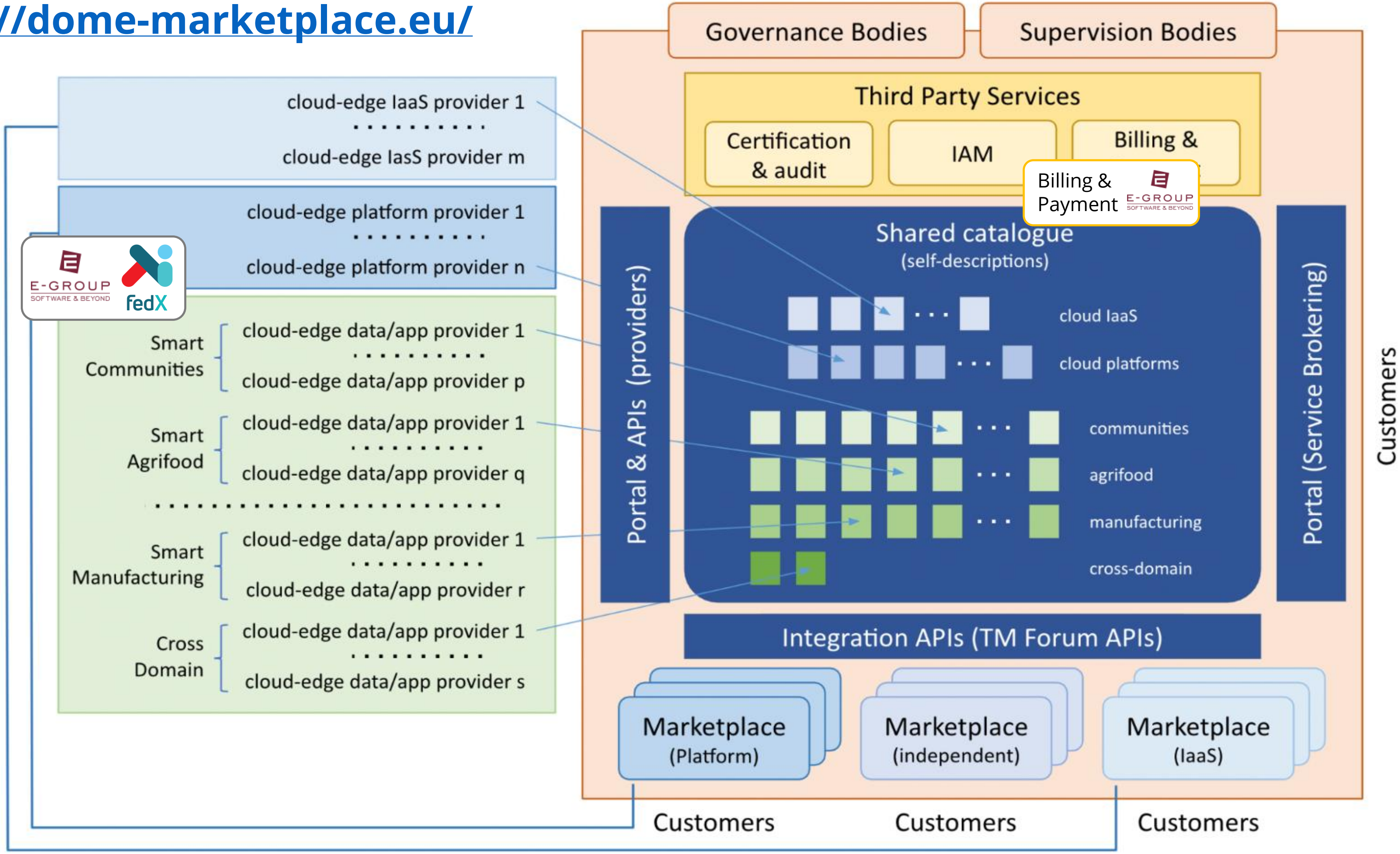
DOME EU Project (E-Group FedX POC)

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

Key EU project https://dome-marketplace.eu/

CLOUD-EDGE PLATFORM PROVIDER

fedX



Governance Bodies
Supervision Bodies
Third Party Services
Certification & audit
IAM
Billing &
Billing & Payment
E-GROUP SOFTWARE & BEYOND
cloud-edge IaaS provider 1
cloud-edge IasS provider m
cloud-edge platform provider 1
cloud-edge platform provider n
E-GROUP SOFTWARE & BEYOND
fedX
Smart Communities
cloud-edge data/app provider 1
cloud-edge data/app provider p
Smart Agrifood
cloud-edge data/app provider 1
cloud-edge data/app provider q
Smart Manufacturing
cloud-edge data/app provider 1
cloud-edge data/app provider r
Cross Domain
cloud-edge data/app provider 1
cloud-edge data/app provider s
Portal & APIs (providers)
Shared catalogue
(self-descriptions)
cloud IaaS
cloud platforms
communities
agrifood
manufacturing
cross-domain
Portal (Service Brokering)
Customers
Integration APIs (TM Forum APIs)
Marketplace (Platform)
Marketplace (independent)
Marketplace (IaaS)
Customers
Customers
Customers

https://www.egroup.hu/dome-the-eus-new-it-cloud-project-with-the-flagship-participation-of-the-hungarian-company-e-group/

1st NATIONAL FEDERATED BIOBANK IN EU

E-GROUP
SOFTWARE & BEYOND

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**.

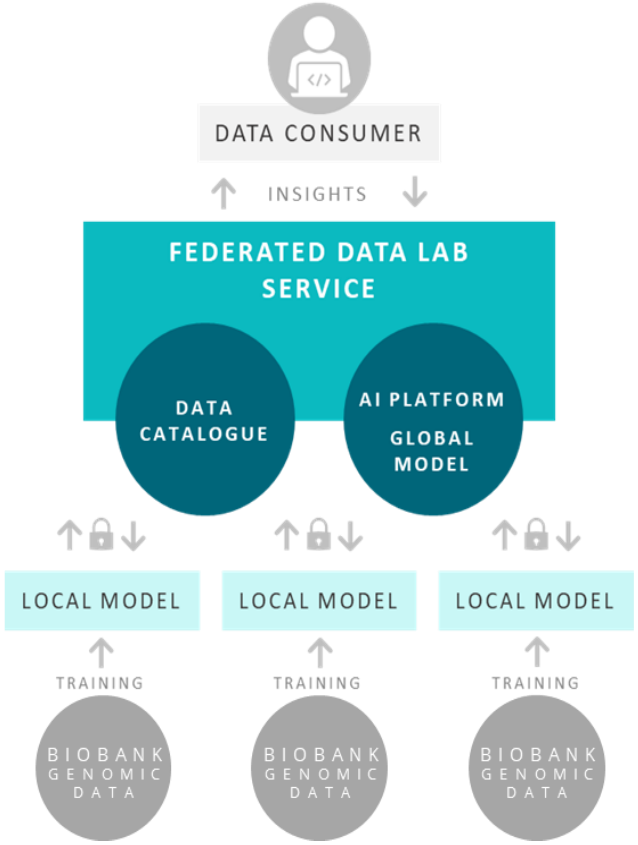
HIGHLIGHTS

- **12 independent biobank i**n 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

MAIN CHARACTERISTICS

TRANSPARENCY | INTEROPERABILITY | FEDERATED DATA EXCHANGE

F. Catalog | F. Cohort Discovery | F. Cohort Statistics | F. Cohort Study



DATA CONSUMER

INSIGHTS

FEDERATED DATA LAB SERVICE

DATA CATALOGUE

AI PLATFORM

GLOBAL MODEL

LOCAL MODEL

LOCAL MODEL

LOCAL MODEL

TRAINING

TRAINING

TRAINING

BIOBANK GENOMIC DATA

BIOBANK GENOMIC DATA

BIOBANK GENOMIC DATA

fedX

Semmelweis University
leading universities of medicine and health sciences

MŰEGYETEM 1782

fedX

E-GROUP
SOFTWARE & BEYOND

Antal KUTHY
E-Group CEO & Founder
antal.kuthy*egroup.hu
(replace * with @)

Ákos TÉNYI PhD
E-Group, Head of SmartData