

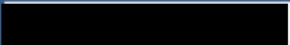
REDACTED DOCUMENT ACCESSIBLE TO THE PUBLIC (01.10.2024)
ONLY MARGINAL PERSONAL DATA HAVE BEEN REDACTED.



LAGO

Lessen Data Access and Governance Obstacles

17 September 2024


Italian National Police
Italy
@LEWP



Funded by the European Union

Grant No 101073951



Call: HORIZON-CL3-2021-FCT-01-04

Topic: *Improved access to fighting crime and terrorism research data*

Type: HORIZON Innovation action

LAGO

Lessen Data Access
and Governance
Obstacles

www.lago-europe.eu



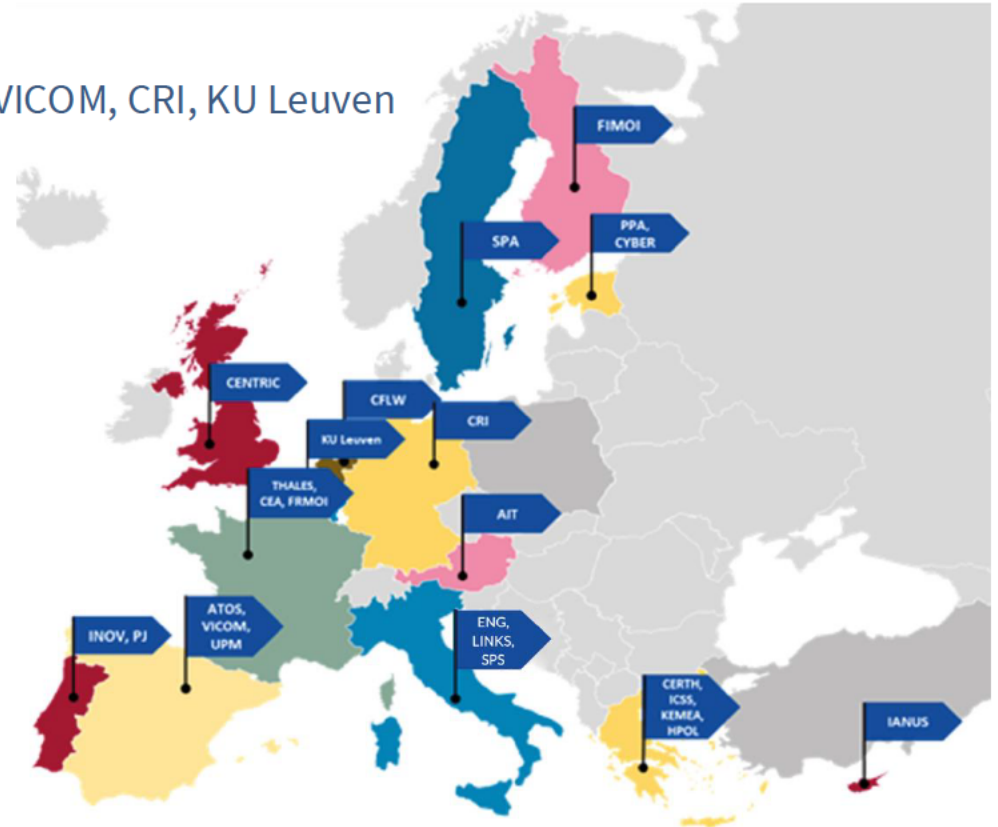
Project details

Consortium: 25 partners from 14 countries

- ❖ **9 Academia and RTOs:** AIT, CEA, CENTRIC, CERTH, ICCS, INOV, LINKS, UPM, VICOM, CRI, KU Leuven
- ❖ **3 Big Industries:** ENG, ATOS, and THALES
- ❖ **2 SME:** CFLW and IANUS
- ❖ **7 Law Enforcement Agencies:** HPOL, FRMOI, PJ, FIMOI, SPA, PPA, SPS
- ❖ **1 Public organisation:** KEMEA

Distinctive elements:

- ❖ **Complementarity in the areas of knowledge and expertise**
- ❖ **Positioning in FCT and Security Research (CERIS, ECSO, GAIA-X)**
- ❖ **Significant numbers of LEAs**
- ❖ **Inclusion of industries** to provide existing processes and solutions



Challenge

LAGO addresses the common and recurring “**Data Issue**” that affects the FCT research landscape, the **lack of domain-specific data** with sufficient quality and quantity **to enable appropriate training and testing of ICT tools and platforms** to support LEAs and security practitioners.



Findings on Barriers to data sharing

Technological Barriers	Organisational Barriers	Economic barriers	Professional & cultural	Policy & Governance
<ul style="list-style-type: none"> • Complexity of technologies • Lack of anonymisation tools • Lack of technological infrastructure for sharing/processing data • Incompatible data formats and lack of interoperability, in adequate data storage and management systems 	<ul style="list-style-type: none"> • Lack of skilled personnel • Lack of collaboration between different stakeholders and competing priorities • Insufficient resources that hinder data sharing. • Lack of clear standards and regulations with organisations to guide data sharing and/or collaborations. 	<ul style="list-style-type: none"> • Inconsistent funding for RDE – related efforts due in part to competing initiatives on a national and EU level. • Costs for tools and datasets. 	<ul style="list-style-type: none"> • Stakeholders with varying professional standards and backgrounds. • Communication gap between security practitioners and technology providers. 	<ul style="list-style-type: none"> • Different legislations in different countries • Lack of applicable policies can lead to disorganisation and mistrust • Lack of clear governance

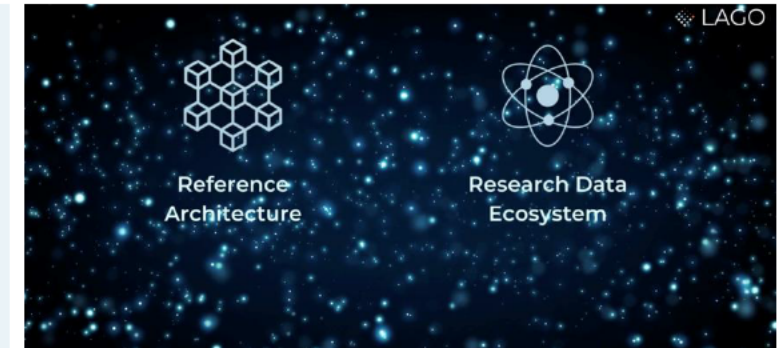
Ambition



LAGO creates a community that brings together **LEAs, policymakers, researchers, industry and practitioners** to improve access to fighting crime and terrorism research data.

Beyond a common repository towards a trusted ecosystem for FCT research data

LAGO aims to go beyond the creation of a common repository for FCT research purpose by **creating the foundations for the sustainable, safe and trusted co-creation, sharing and maintenance** of representative datasets.



Scientific and technological innovation

- ❖ Pillar 1
Data Creation
- ❖ Pillar 2
Data Usage

- ❖ Pillar 3
Legal, Ethical and EU policies
- ❖ Pillar 4
Trust and Governance Framework

Use Cases

Uses cases

- ❖ Use Case 1
Fight against illicit trafficking of arms
- ❖ Use Case 2
Biometric data exchange in criminal cases
- ❖ Use Case 3
Public space protection through video analytics
- ❖ Use Case 4
Deepfake detection in photos and movies
- ❖ Use Case 4
Edge computing for public safety
- ❖ Use Case 5
Counter Terrorism
- ❖ Use Case 6
Illicit goods trafficking
- ❖ Use Case 7
Training for research purpose

Enablers to data sharing

- ❖ **Increase awareness and understanding on potential benefits of a RDE** among relevant stakeholders
- ❖ **Common standards and protocols** to facilitate data sharing by enabling interoperability and reducing technical barriers
- ❖ **Policy and regulatory frameworks** that support data sharing and collaboration while protecting data ownership, privacy and security
- ❖ **Training and education programs** to promote best data sharing practices, build data management skills and raise awareness of benefits of data sharing
- ❖ **Collaborative partnerships** among different stakeholders to build trust, share expertise and resources

The LAGO principles



Decentralization: FCT research data and datasets are not centralized, but created, provided, and made available by data providers to users in a federated environment.



Data sovereignty: ownership and control of data are retained by providers/owners.



Security and trust: the RDE maintains confidence in the identity and capability participants and provides the measures and tools to protect the integrity and security of data and operations on them.



Data quality: FCT research requires high-quality datasets to train and test data-driven and AI/ML prototypes and solutions.



Openness: rules, specifications, and protocols for data sharing in the context of the RDE are open.



Transparency: data sharing and exchange are transparent, tracked, traceable, and accountable.



Proportionality and risk: the RDE ensures and provides capabilities for assessing the risk and proportionality between the lawfulness and possible interference with the fundamental rights in accessing and providing data and datasets.



Interoperability: several systems or services, both providers and users, will be enabled to exchange and properly use harmonized (in format, structure, and semantics) research data within the RDE.



Portability: data is described in a standardized protocol that enables transfer and processing to increase its usefulness as a strategic resource.



Ethics, Legal and Privacy: FCT research data normally include personal and sensitive data; the RDE will consider and comply with legal and ethical rules of operation (including privacy and data protection) and fundamental rights as well as applicable legislations in EU Member States.

Challenge: Risks of sharing datasets

- ❖ When sharing FCT research data, a number of **risks** may materialize, especially in cases of **unauthorized** data disclosure or modification, due to various **factors**:

Technological factors

(e.g., poor data anonymization, data transfer, data storage)

People-related factors

(e.g., poor training or experience in handling sensitive data)

Institutional factors

(e.g., related to cybersecurity protocols, data governance framework)

Legal factors

(e.g., lack of comprehensive license agreement)

- ❖ Consequently, data providers can be very reluctant to share datasets

- ❖ Possible solution to help **lift** these barriers:

A **risk assessment methodology** to measure, evaluate and help mitigate risks for granting access to FCT research data



Data Quality challenges



- ❖ A common and recurring challenge in the FCT landscape is the **lack of domain-specific data** with sufficient **quality**

The quality of the training and testing data (including the quality of their structure and labelling, and how well these data represent the problem to be tackled) affects the accuracy of data-driven digital tools

Having a scientifically satisfactory amount of up-to-date high-quality and realistic data is key to producing useful tools in support of law enforcement

- ❖ Several factors that can contribute to the inadequacy of data quality



Diverse Data Sources



Lack of Standardization



Human Error Bias



Data Privacy Concerns



Data Volume Velocity

- ❖ **Possible Solution** to help lift these barriers:

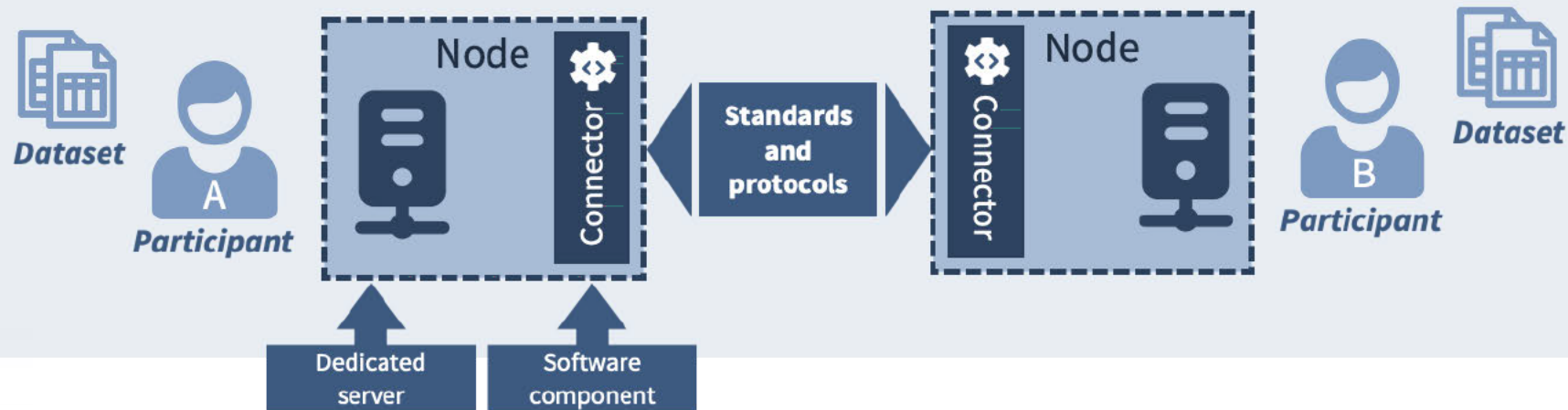
A **Data Quality Assessment framework** to assist with the verification of the data used in training and testing, supporting a range of different indicators (data age, completeness, variety, biases, etc.) and providing insights regarding potential limitations



RDE overview

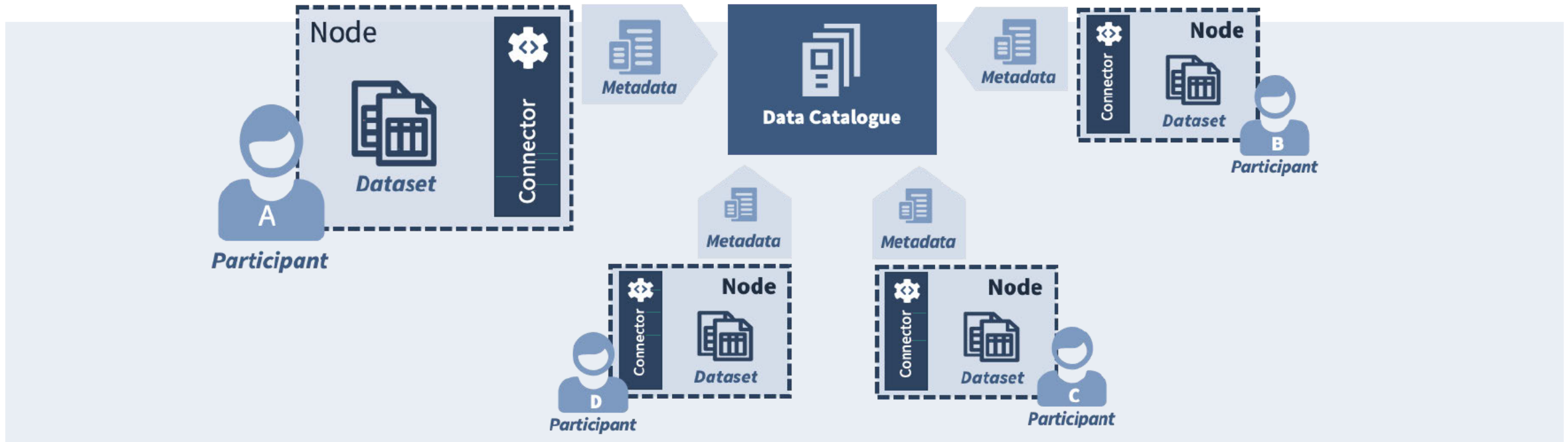
The **Research Data Ecosystem** (RDE) is made of

- ❖ **Participants**, actors that share or need access to data
- ❖ **Research Datasets**
- ❖ **Standards** and **Protocols** for the interoperable exchange of datasets
- ❖ **Technical components** to implement the RDE standard protocols for allowing participants to access research datasets



Enabling access to research data

- ❖ To make participants aware of a dataset available for sharing, the data provider can register the dataset on the **Data Catalogue**
- ❖ To ensure **data sovereignty**, only **metadata** about datasets are published on the Data Catalogue, while data remain **safely stored on participant premises**



Metadata include information about the **nature** of the datasets, the **usage policies** and **licenses** under which the dataset is made accessible





Questions?


Italian National Police