



# MOSAICrOWN

Multi-Owner data Sharing for Analytics and Integration  
respecting Confidentiality and OWNeR control

## Introduction to MOSAICrOWN

Rigo Wenning

European Research Consortium for Informatics and  
Mathematics (ERCIM)

The ecosystems of HPC and Big Data/AI: what can they learn from  
each other? - March 15, 2022



- Data produced all over the world are growing rapidly
- The availability and sharing of data are essential for the economy and society
- European data strategy

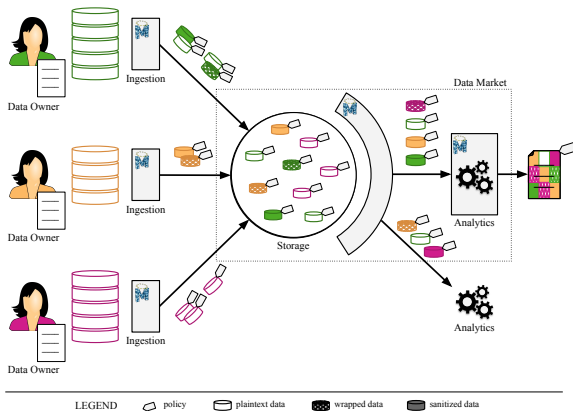
But ....

- Data-driven innovations are hampered by the
  - lack of technology for the construction of trusted and secure platforms where personal and/or company-confidential data can be securely stored, shared, and processed

# MOSAICrOWN's goal and vision



Enable data sharing and collaborative analytics in multi-owner scenarios, while ensuring proper protection of sensitive or company-confidential information





1. Support a rich set of protection requirements
2. Provide a data governance framework
3. Provide efficient and effective techniques for data wrapping
4. Provide efficient and effective techniques for data sanitization
5. Provide effective exploitation in real operational environments, enabling demonstration of the applicability and flexibility of the project's innovations and actual impact



- **Personal data protection is improved**, and compliance with regulations (e.g., GDPR) is made easier
- **Citizens' trust is improved**, by privacy-aware transparency and control features across data platforms
- **Value-creation from data is improved**, as barriers to data market adoption are reduced
- **New data market opportunities are enabled**, supporting European digital economy and contributing to the European Green Deal



MOSAICrOWN identifies three dimensions characterizing the problems and challenges

- **Requirements capturing and representation**  
policies regulating access, sharing, usage and processing
- **Enforcing technologies**  
data wrapping / sanitization
- **Enforcement phase**  
ingestion / storage / analytics

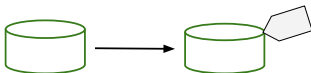


Data owners need to have a way to express their requirements and having them enforced



Data owners need to have a way to express their requirements and having them enforced

- **Policies** regulate access, sharing, usage and processing of data







## Techniques and mechanisms for enforcing data protection





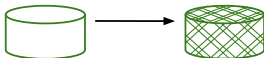
## Techniques and mechanisms for enforcing data protection

- **Wrapping**: provide protection by (partially or completely) disabling visibility of data while preserving some functionality



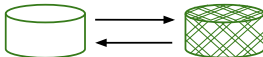
## Techniques and mechanisms for enforcing data protection

- **Wrapping**: provide protection by (partially or completely) disabling visibility of data while preserving some functionality



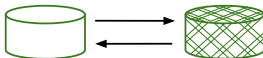
## Techniques and mechanisms for enforcing data protection

- **Wrapping**: provide protection by (partially or completely) disabling visibility of data while preserving some functionality



## Techniques and mechanisms for enforcing data protection

- **Wrapping**: provide protection by (partially or completely) disabling visibility of data while preserving some functionality

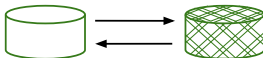


- **Sanitization**: provide protection by returning an obfuscated (e.g., not precise) version of the data

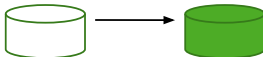


## Techniques and mechanisms for enforcing data protection

- **Wrapping**: provide protection by (partially or completely) disabling visibility of data while preserving some functionality

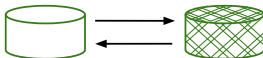


- **Sanitization**: provide protection by returning an obfuscated (e.g., not precise) version of the data

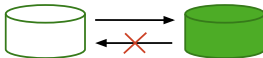


## Techniques and mechanisms for enforcing data protection

- **Wrapping**: provide protection by (partially or completely) disabling visibility of data while preserving some functionality

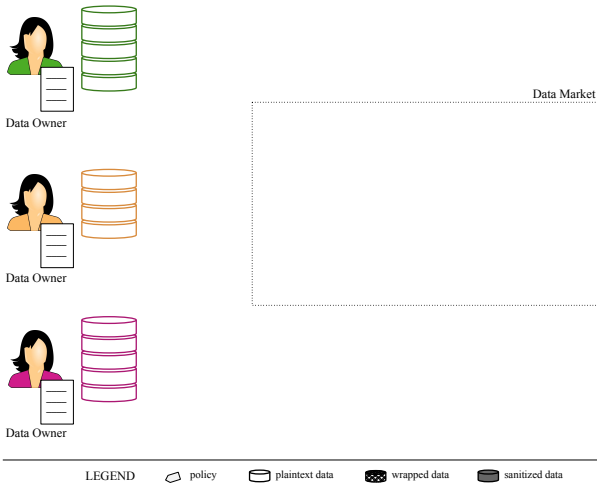


- **Sanitization**: provide protection by returning an obfuscated (e.g., not precise) version of the data



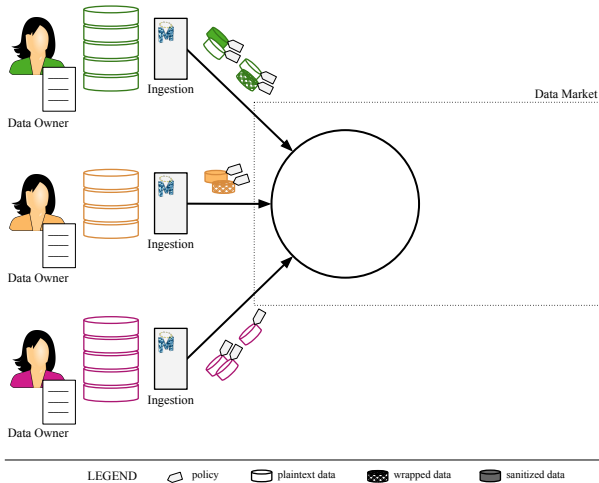


- Ingestion / Storage / Analytics

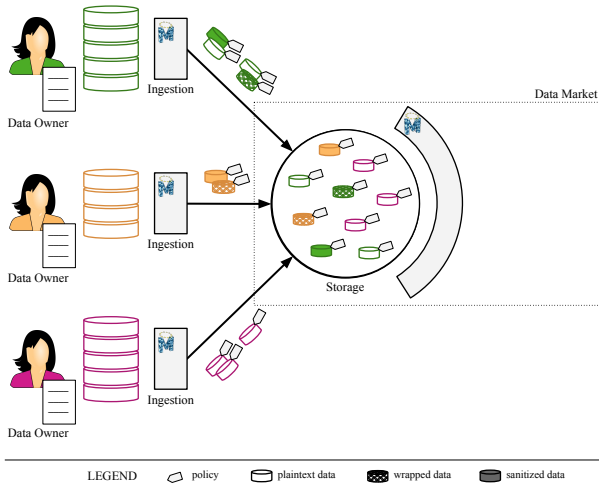




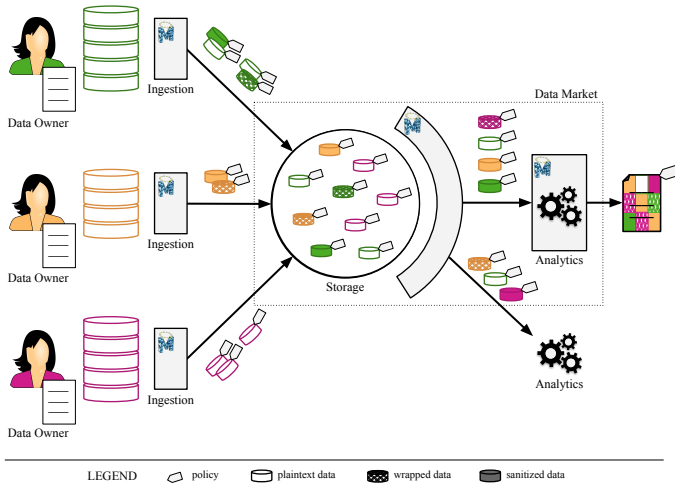
- Ingestion / Storage / Analytics



- Ingestion / Storage / Analytics



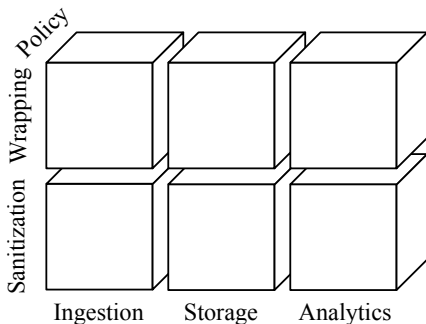
- Ingestion / Storage / Analytics



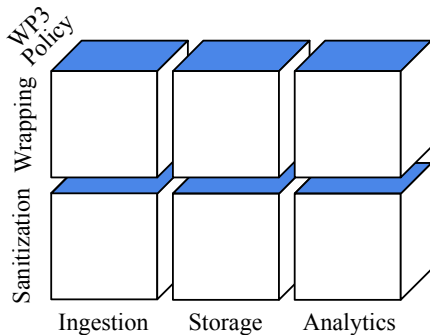


- Tackle challenges **gradually**, providing **modular** techniques and tools
- Structure the work along the different dimensions, providing a data governance framework for specifying **policies** together with data **wrapping** and **sanitization** techniques protecting data in the whole life cycle (**ingestion**, **storage**, and **analytics**)
- Ensure correspondence with use cases, with common aspects but also particular focus on different aspects/dimensions

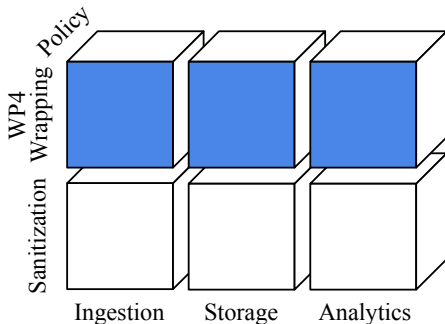
# Dimensions of the problem



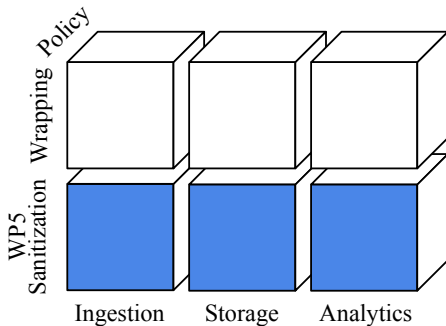
- Requirements capturing and representation (policy)
- Enforcing technologies (data wrapping, data sanitization)
- Enforcement phase (ingestion, storage, analytics)



- Reference metadata model
- Policy model and language
- Policy management



- Basic techniques
- Enforcement mechanisms
- Controlled query execution
- Incentives and economic aspects

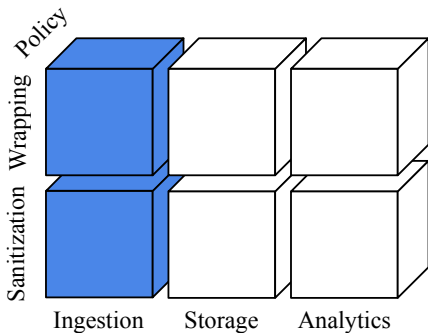


- Privacy metrics and risks
- Data sanitization techniques
- Collaborative computation for sanitization

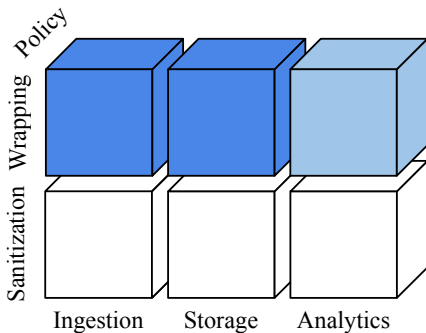




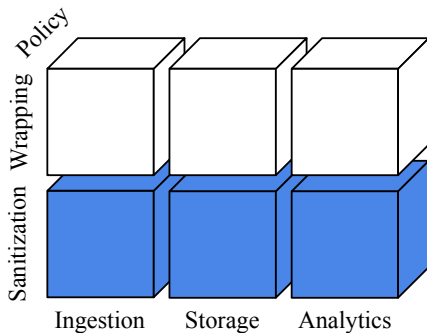
- **Real-world** problems for which there are **open** privacy / confidentiality problems that cannot be addressed by current technology
- Different **application domains**:
  - Intelligent connected vehicles (UC1, EISI)
  - Data market for financial data (UC2, MC)
  - Cloud-based market for consumer analytics (UC3, SAP SE)
- Focus on different **aspects** / **dimensions**



- Data exchange platform for the automotive industry
- Governance framework for managing collected data
- Protect sensitive portion of data injected in the data market



- Financial transactions data from banks and organizations
- Selective and privacy-aware processing within the data market
- Privacy-aware analytics outside the data market



- Cloud services for SAP SE customers
- Selective sharing of anonymized data
- Privacy-preserving data analytics



- Comprehensive [method to gather requirements](#)
- [Data governance framework](#) for modeling data, metadata, and enabling protection requirement specifications
- [Policy model, language](#), and accompanying [enforcement engine](#) for protecting data throughout the complete life-cycle
- [Data wrapping solutions](#), providing for a (reversible) self-enforcing layer of protection to data in storage and processing
- [Data sanitization solutions](#), enabling efficient and effective privacy-aware data release and analytics



# Thank you!



UNIVERSITÀ  
DEGLI STUDI  
DI MILANO

DELL EMC



UNIVERSITÀ  
DEGLI STUDI  
DI BERGAMO



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 825333.