



Unlocking the potential of digital data

Privacy Preserving Data Sharing

Vincent Naessens

DistrINet

Bad News: closing event TeTra Start2AIM



Good News: New Tetra project Approved!

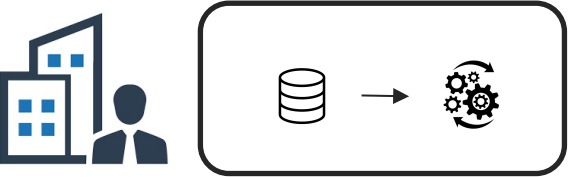
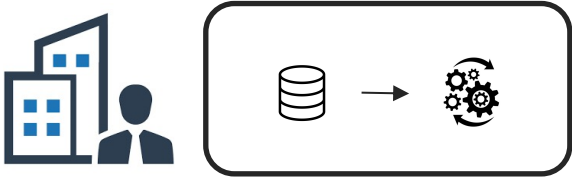
- › Title: Privacy Preserving Data Sharing (DASH)



- › Kick-off meeting: 22 02 2022 @ 16h



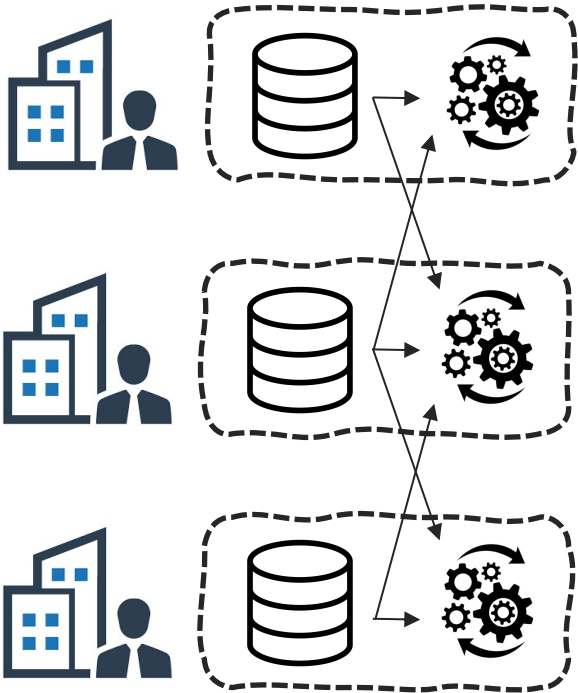
Digital Data is the new gold





- › **Digitization** (*first wave*)
 - › Personnel management
 - › Customer data
 - › Maintaining inventories

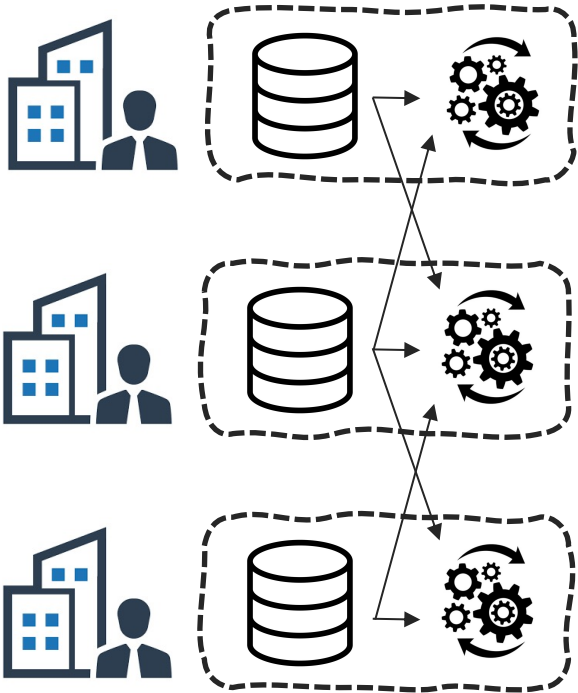
- › **Advanced decision making** (*second wave*)
 - › recommendations
 - › Predictions
 - › Strategic decisions

Improving business intelligence



- › **Increasing data collection** 
 - › Fine-grained data collection
 - › Integrating external data sources
 - › *Increasing storage capacities*
- › **Increasing processing power** 
 - › Machine learning and AI technology
 - › Optimization algorithms
 - › *Increasing computing power*

Some application domains



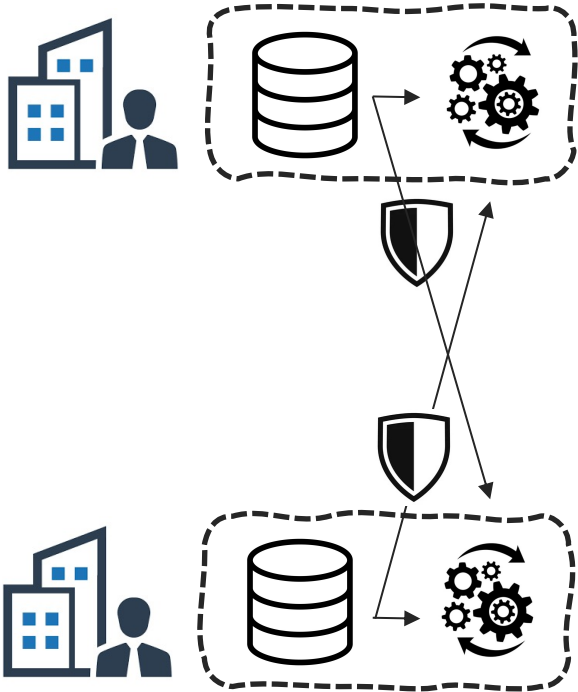
› **Crime control**

- › Goal: optimal allocation of police forces
- › Combining governmental and financial data
- › Personal + company data

› **Health, activity and lifestyle**

- › Goal: improving lifestyle
- › Health, food and activity tracking
- › Sensitive personal data

Controlled release of sensitive data

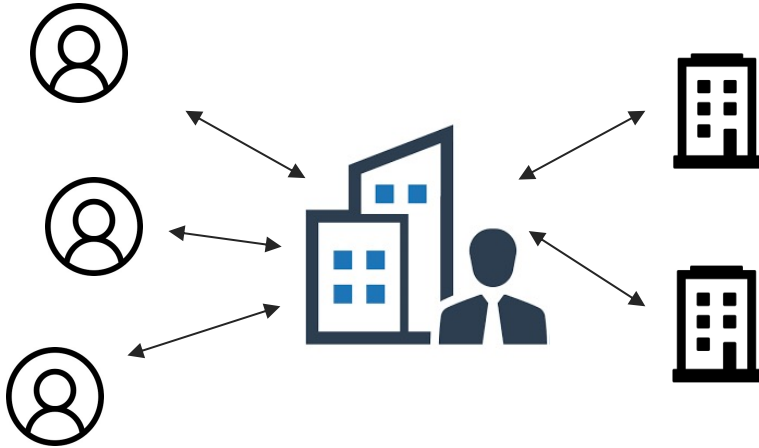


› Why *controlled* release?

- › Compliance with privacy regulation
- › Discrimination
- › Reputation damage
- › Economic loss

› How *controlled* release?

Techniques for controlled release



> **User control**

- » Data minimization
- » Local differential privacy

> **Controlled query handling (pull)**

- » Query perturbation
- » Restricted query handling
- » *Differential privacy* → *privacy budget*

> **Controlled dataset transfer (push)**

Controlled dataset transfer



- › **Pseudonymization** ↔
 - › Replacing fields with pseudonyms
 - › Reversible
- › **Anonymization** ✂
 - › Stripping elements
 - › Irreversible

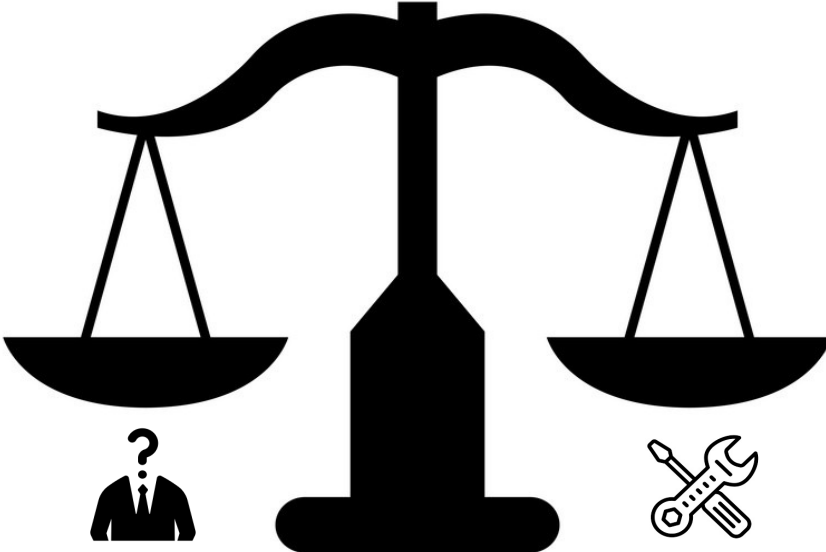
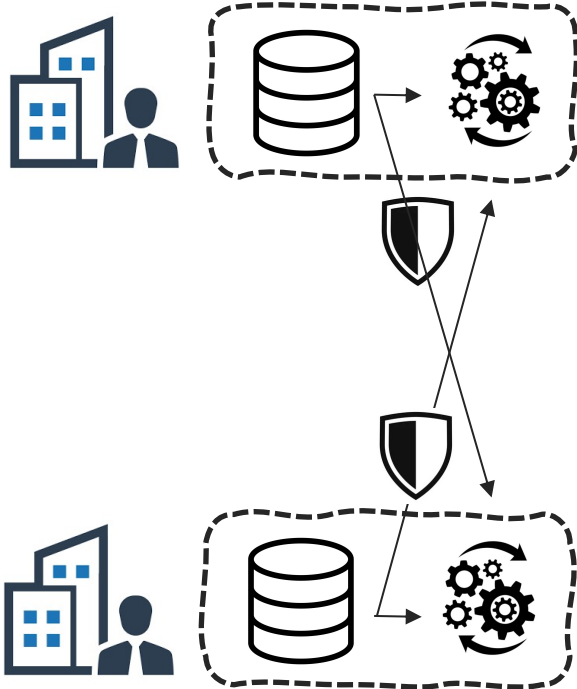


Emerging challenges

- › The privacy ↔ utility balance
- › Outsourcing
- › Evolving attack(er)s
- › Every increasing complexity

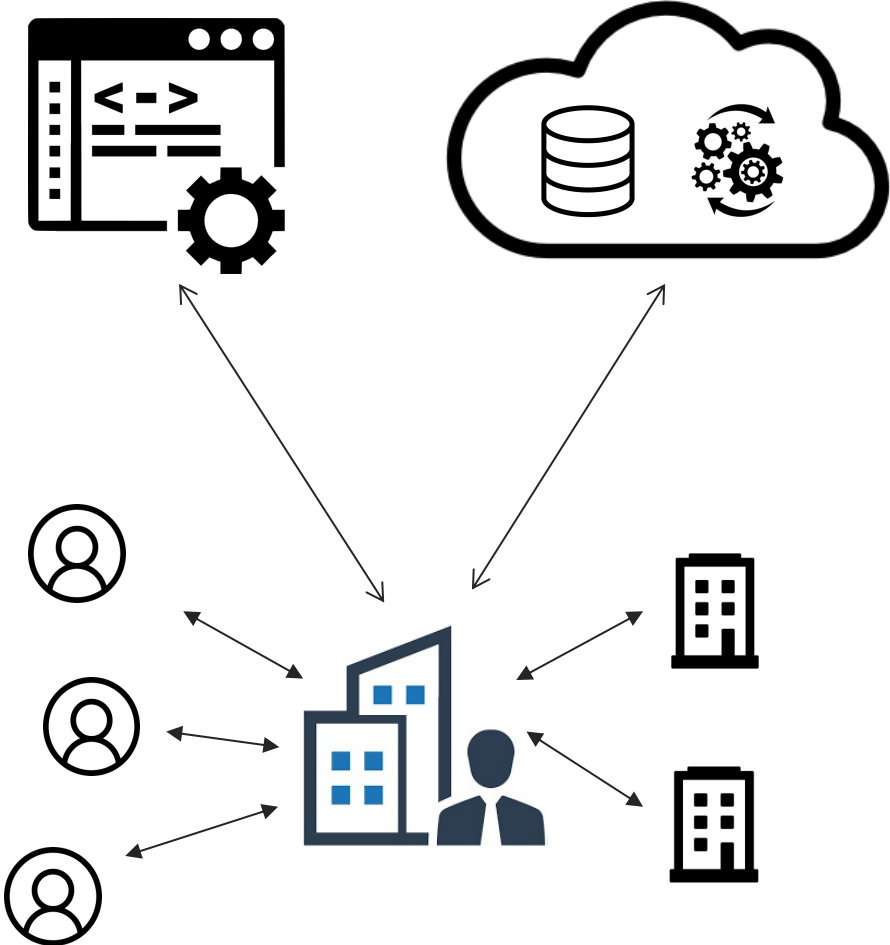
- › **The privacy ⇔ utility balance**
- › Outsourcing
- › Evolving attack(er)s
- › Every increasing complexity

The privacy ↔ utility balance

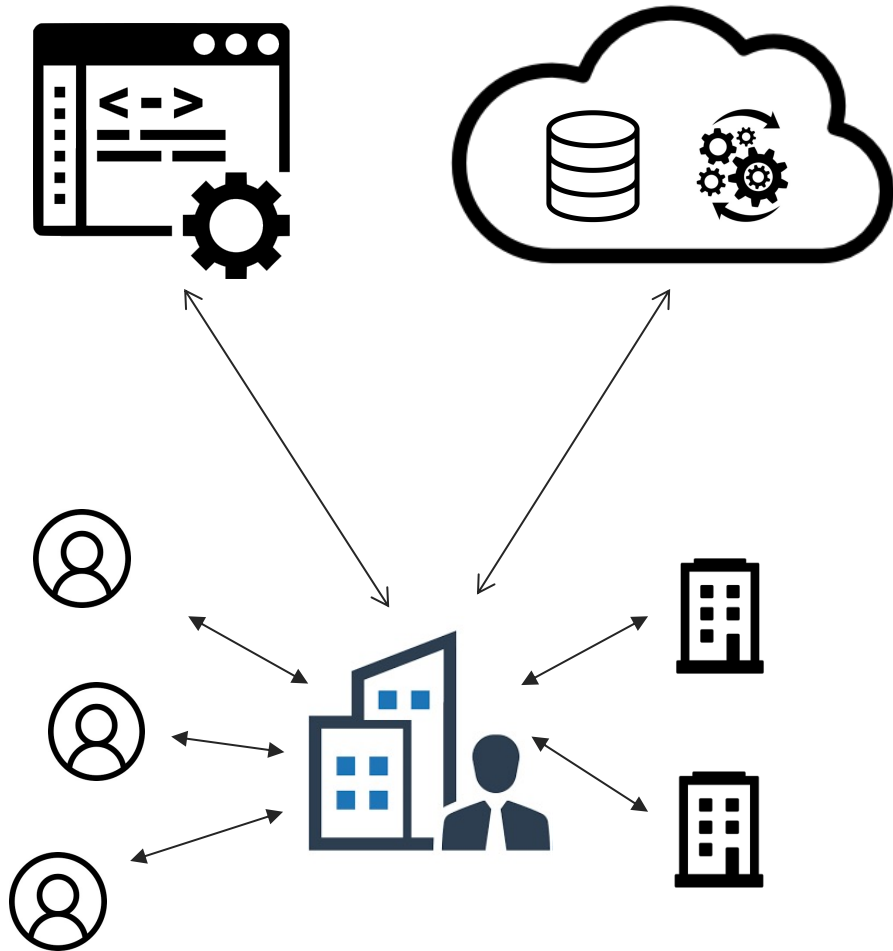


- › The privacy ↔ utility balance
- › **Outsourcing**
- › Evolving attack(er)s
- › Every increasing complexity

Outsourcing



Outsourcing



Trusted Third Party



Honest-but-Curious Service Provider

- › The privacy ↔ utility balance
- › Outsourcing
- › **Evolving attack(er)s**
- › Every increasing complexity

Evolving Attacks

› **Attack vectors**

- › Data in Transit → secure communication channels
- › Data in Rest
- › Data during computation

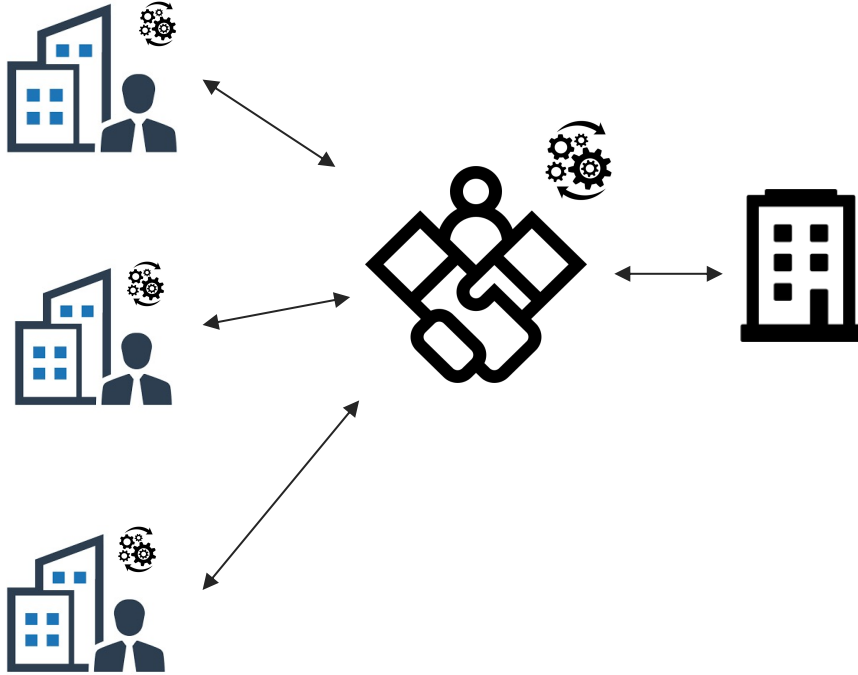
› **Attacks on publicly available datasets**

- › The Prosecutor → targeting a specific individual in dataset
- › The Journalist → targeting any individual
- › The Marketeer → re-identifying a large number of IDs



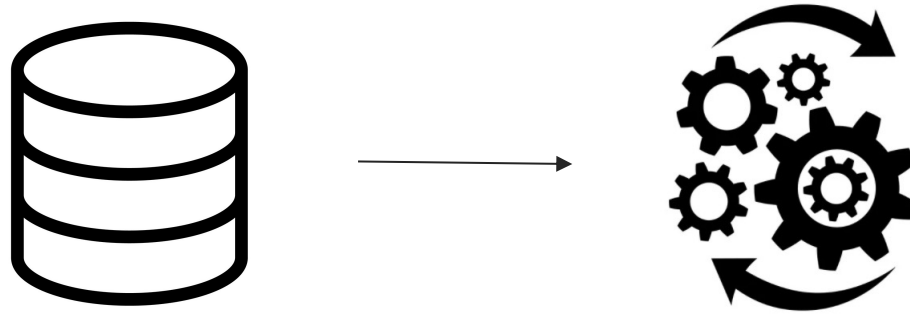
- › The privacy ↔ utility balance
- › Outsourcing
- › Evolving attack(er)s
- › **Every increasing complexity**

Multiple data controllers (1/2)



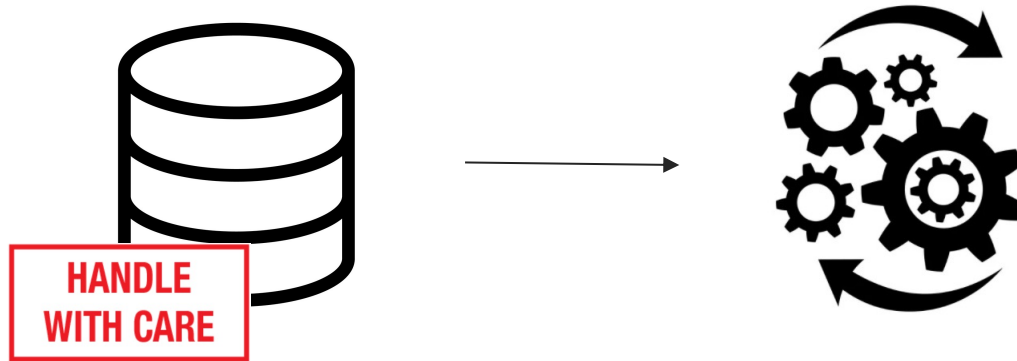
- › **Federated computing**
 - › Data controllers release properties
 - › Leakage in case of multiple queries
- › **(Fully) homomorphic encryption**
 - › Encrypted processing
 - › Static set-ups / simple operation

Conclusions



No processing without data

Conclusions



No processing without data

Conclusions

- › Protection is important during the **whole data lifecycle**
 - › collection – storage – processing - release
 - › privacy-by-design
- › Apply realistic **trust** assumptions
 - › Who are my allies?
 - › Honest-but-curious third parties
- › Embrace **innovative software technologies**
 - › Statistical methods - AI - ML
 - › Cryptographic technologies



You are welcome!

- › Title: Privacy Preserving Data Sharing (DASH)



- › Kick-off meeting: 22 02 2022 @ 16h

