

# Pseudonymisation As A Service

and Data Protection Assisted by Computers

Cédric Lauradoux

8 décembre 2021

# Why controllers fail with pseudonymisation ?

## ▶ **Fundamental pseudonymisation problems :**

- ① Controllers don't know what to protect.
- ② Controllers don't know which function to use.  
— People are not trained to data protection.

## ▶ **Can we solve these two issues ?**

- Use programs to identify sensitive data.
- Delegate/outsource pseudonymisation.

# Importance of personal data identification

- ▶ **Fundamental task for Data Privacy Officers :**
  - Privacy Impact Assessment
  - Data Protection Enforcement
- ▶ **Difficult task for DPOs :**
  - Easy to underestimate data sensitivity.
  - Weak spot of the security foundations
- ▶ **Can computers assist DPOs ?**

# How to identify personal data

## ▶ Main obstacles to automatisation :

### ① Machine learning

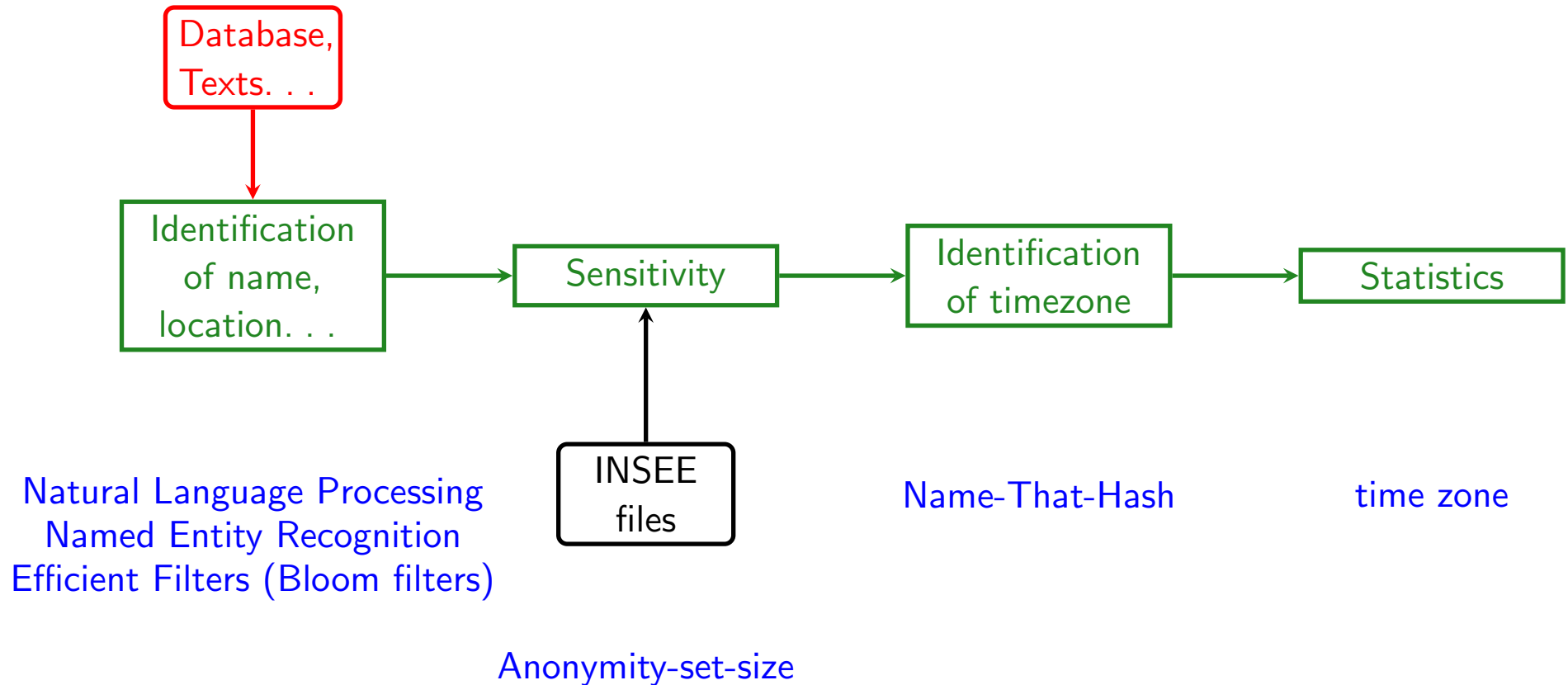
- Filtering with efficient lists
- Named Entity Recognition (NER)

### ② Pseudonym identification

- Entropy (behave to random values)
- Name-That-Hash

### ③ Statistics

# Our Detector Architecture



- ▶ **There are some limits** : sensitivity is contextual.

# Local Pseudonymisation

- ▶ Data controllers have their pseudonymisation scheme  
They need to use secure schemes.
- ▶ **People need to be trained.**  
People need to make the good choices to avoid mistakes and data breaches.
- ▶ **Alternative** : call an expert and outsource pseudonymisation.

# Outsourced Pseudonymisation

## Advantages

- ▶ **Training** : nobody need to be trained.  
You need to know an expert.
- ▶ **Up-to-date** : responsibility of the expert.
- ▶ **Limited responsibility** : you can blame the expert.
- ▶ It is not my fault :

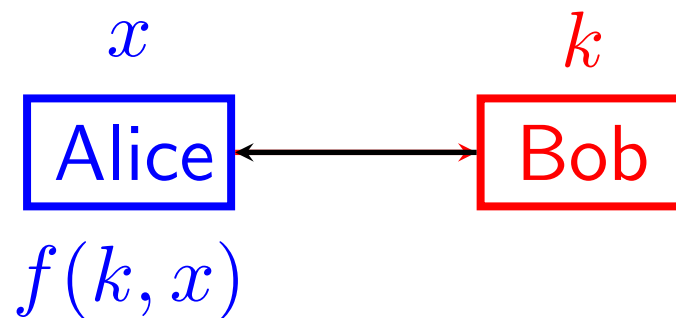
# Problem with Delegation ?

## Problems

- ▶ There are legal implications !  
The entity which pseudonymises is a data processor.
- ▶ The data must be transferred. . . (secure channel)
- ▶ We centralised personal data on a single server.  
Risk of data breaches !
- ▶ **Cryptography to the rescue : OPRF.**

# Oblivious Pseudo Random Functions

- ▶ An **Oblivious Pseudorandom function** (OPRF) is a protocol that allows a party Alice to securely compute a **pseudorandom function**  $f(k, x)$  on an input  $x$  contributed by Alice and on the key  $k$  contributed by Bob, such that :
  - Alice doesn't learn  $k$
  - Bob learns nothing from the interaction



# Conclusion

- ▶ **Computers can assist DPOs and DPAs !**  
We can automatise painful tasks (aka audit)  
DPOs and DPAs can focus on more important topics.
- ▶ We need **Data Protection Assisted by Computer** :
  - Personal data detector (for PIA, . . . )
  - Pseudonymisation As A Service
- ▶ Otherwise, the GDPR cannot be fully enforced.  
We need to provide tools to DPOs and DPAs.