International Conference PACBB

How can photo sharing inspire sharing genomes?

Vinicius V. Cogo[1], Alysson Bessani[1], Francisco M. Couto[1]
Margarida Gama-Carvalho[2]
Maria Fernandes[3], Paulo Esteves-Verissimo[3]

[1]LASIGE, University of Lisbon
[2]BioSys, University of Lisbon
[3]SnT, University of Luxembourg

U LISBOA | UNIVERSIDADE DE LISBOA

uni.lu UNIVERSITÉ DU LUXEMBOURG

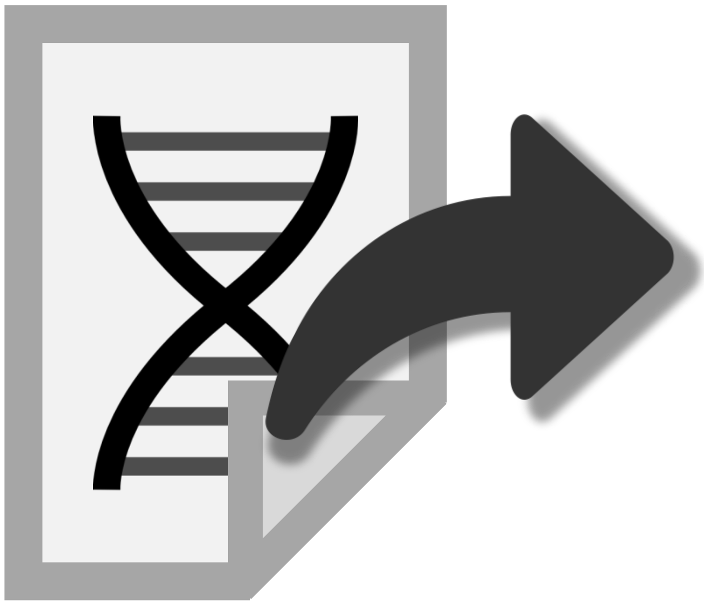
PACBB'17. June 21-23, 2017. Porto, PT.

Context

People usually **are aware** of the **privacy risks** of **publishing photos** online …

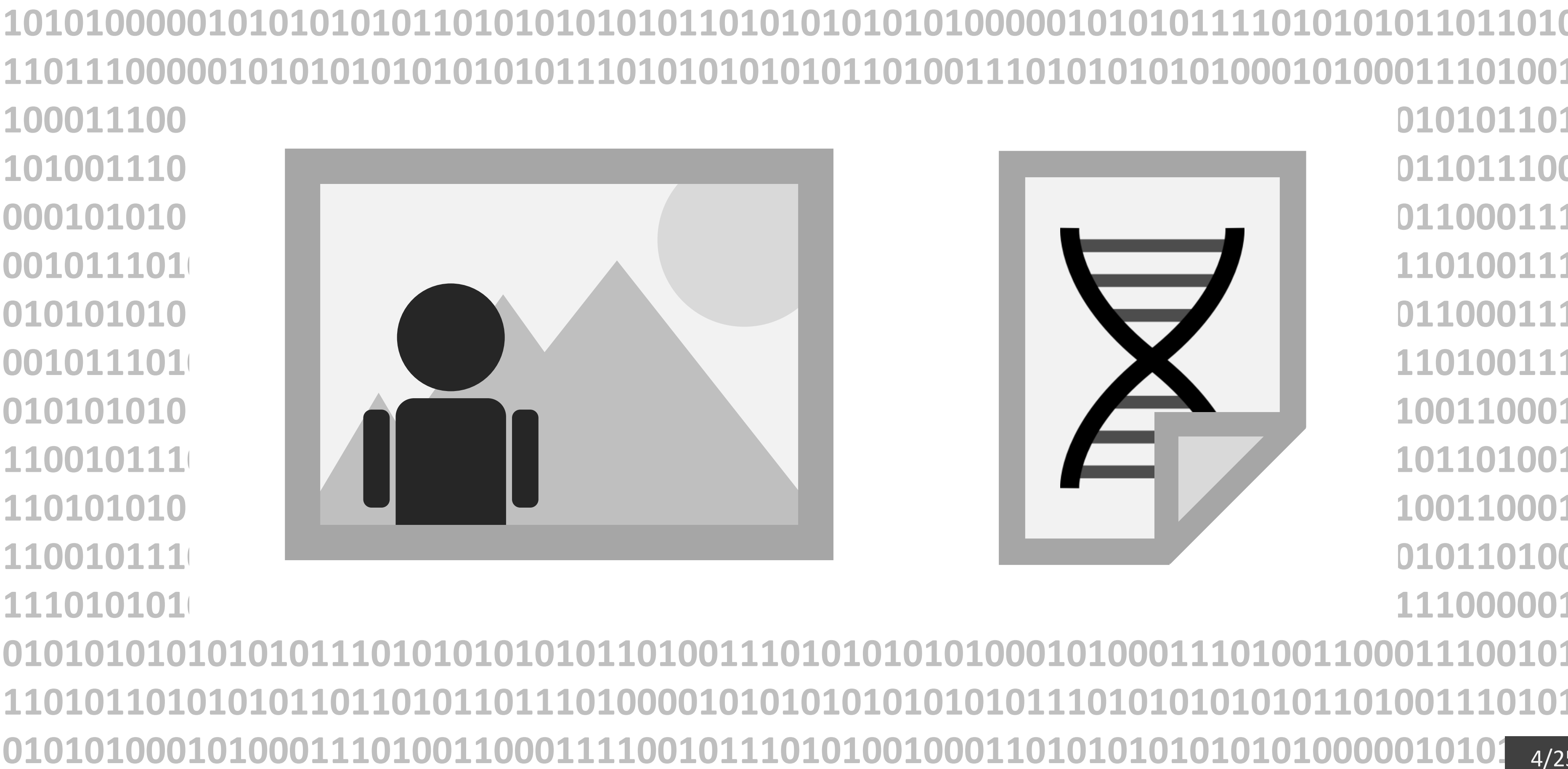
Context

… but **these risks** are **less evident** when
sharing human ***genomes***



Context

Modern **photos** and sequenced **genomes** are both **digital representations** of real lives



Goals

1. Analogy

Present an analogy between the privacy aspects of sharing photos and sharing genomes, which clarifies the privacy risks in the latter to the general public.

2. Informed model

Illustrate an alternative informed model to share genomic data according to the privacy-sensitivity level of each portion.

Outline

Context

1. Analogy

2. Informed model

Final remarks

Outline

Context

1. Analogy

2. Informed model

Final remarks

1. Analogy – sharing photos and sharing genomes

Overview

Goal:
Present an analogy between the privacy aspects of sharing photos and sharing genomes, which clarifies the privacy risks in the latter to the general public.

Method:
Describe the mostly-known privacy risks in sharing photos and then make the analogy to sharing genomes

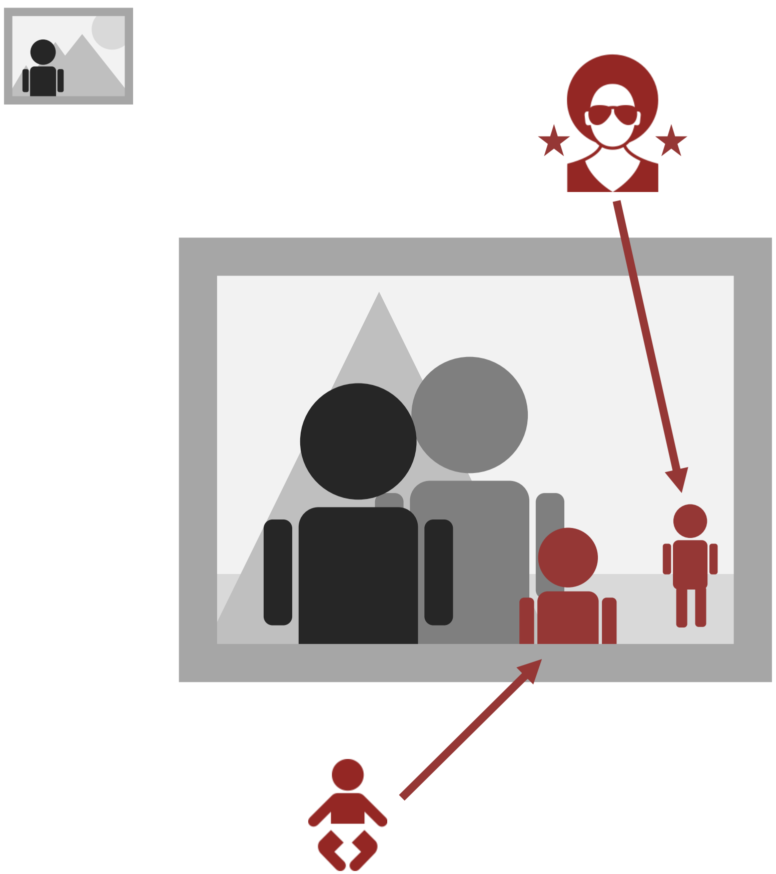
Aspects:
Special data portions; Privacy of others;
Detection; Classification; Impact of sharing

1. Analogy – sharing photos and sharing genomes

Some **portions** of data are **more privacy-sensitive** than others

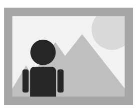
1. Analogy – sharing photos and sharing genomes

One's data may affect
the **privacy of others**



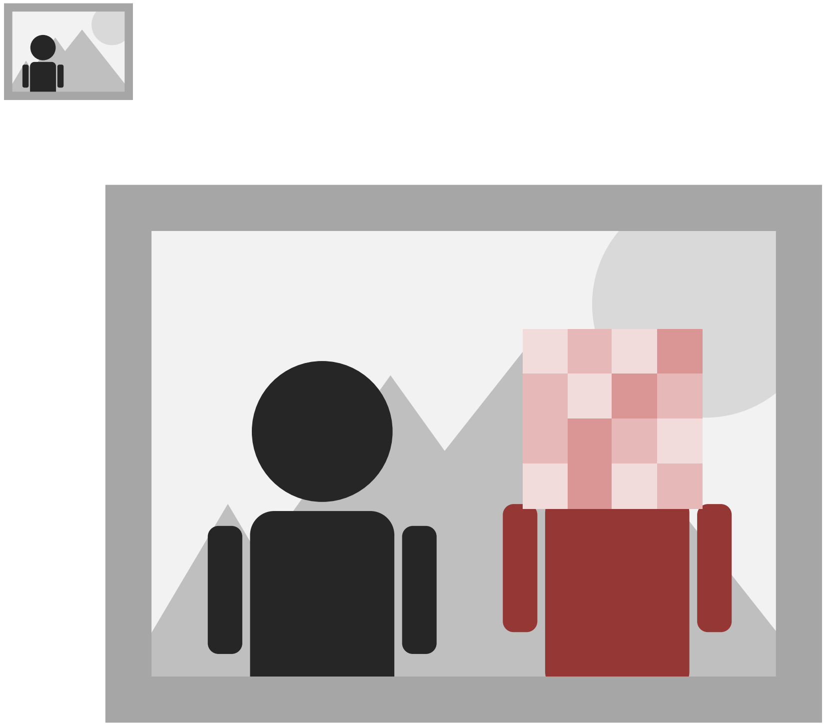
1. Analogy – sharing photos and sharing genomes

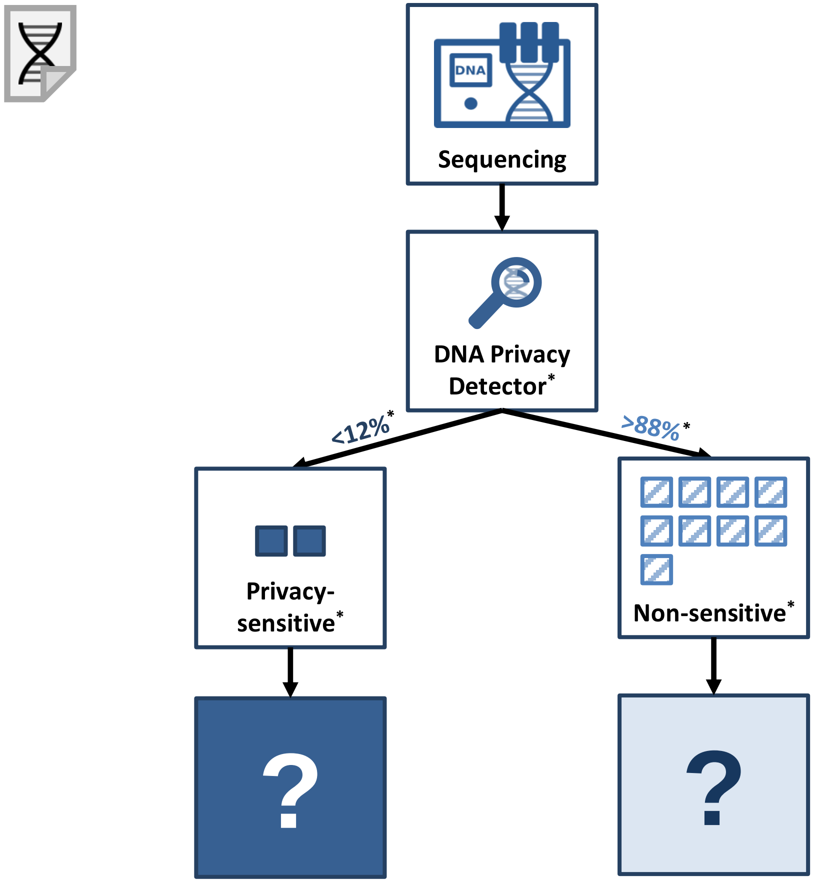
Systematically detecting the privacy-sensitive portions of data is **feasible**



1. Analogy – sharing photos and sharing genomes

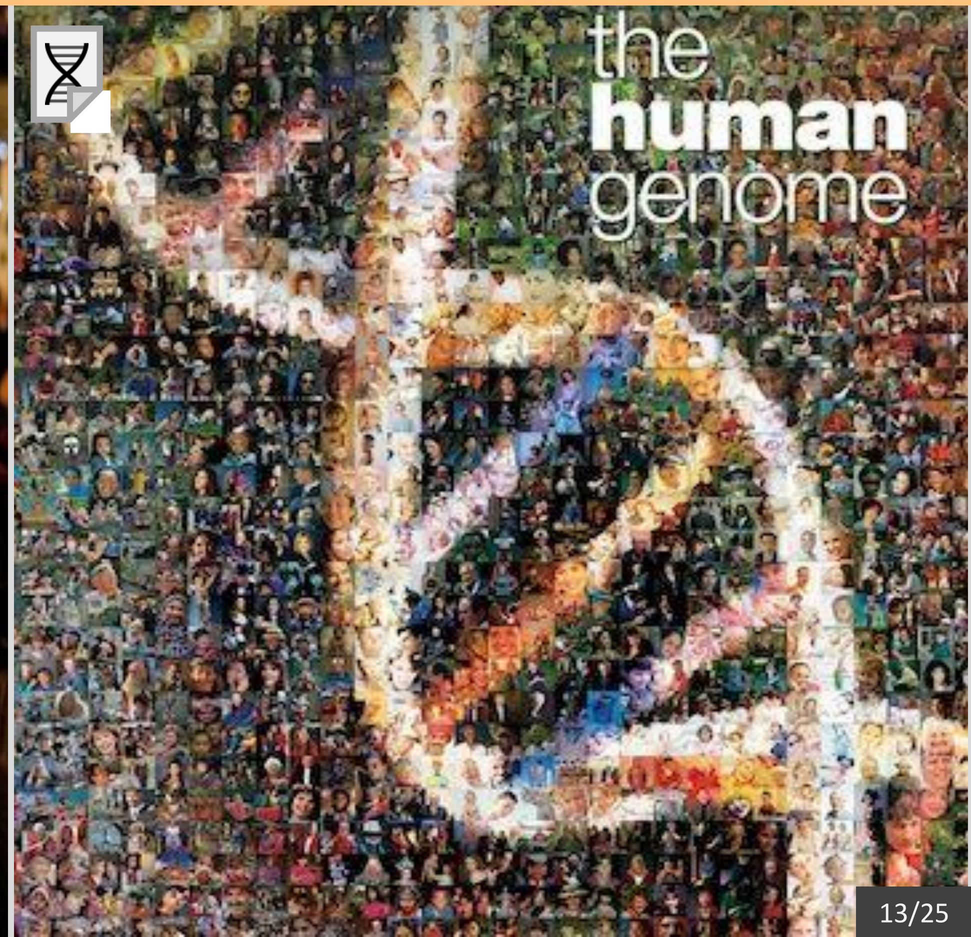
After classifying the portions,
<u>decide how to share</u> them





1. Analogy – sharing photos and sharing genomes

The **impact** of data sharing
is **unpredictable**. **Every** **+1** **matters**.



Outline

Context

1. Analogy

2. Informed model

Final remarks

2. Informed model – sharing genomes

Overview

Goal:

Illustrate an alternative informed model to share genomic data according to the privacy-sensitivity level of each portion.

Method:

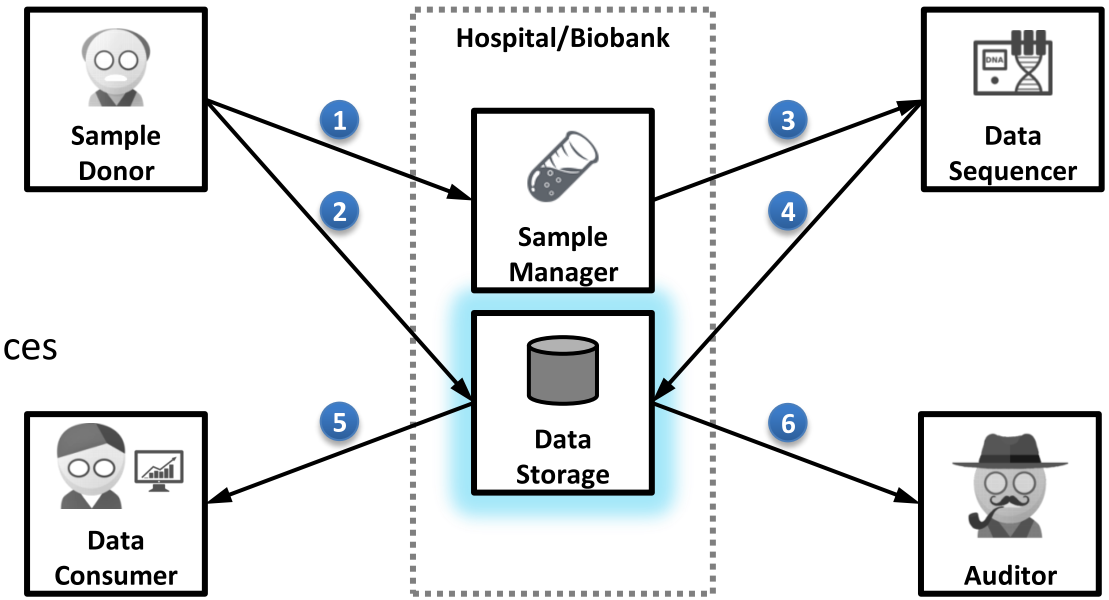
Describe the players and their interactions with the model; describe some important internal components

Important internal components:

Evolution module; Storage; Access control; Auditability

2. Informed model – sharing genomes

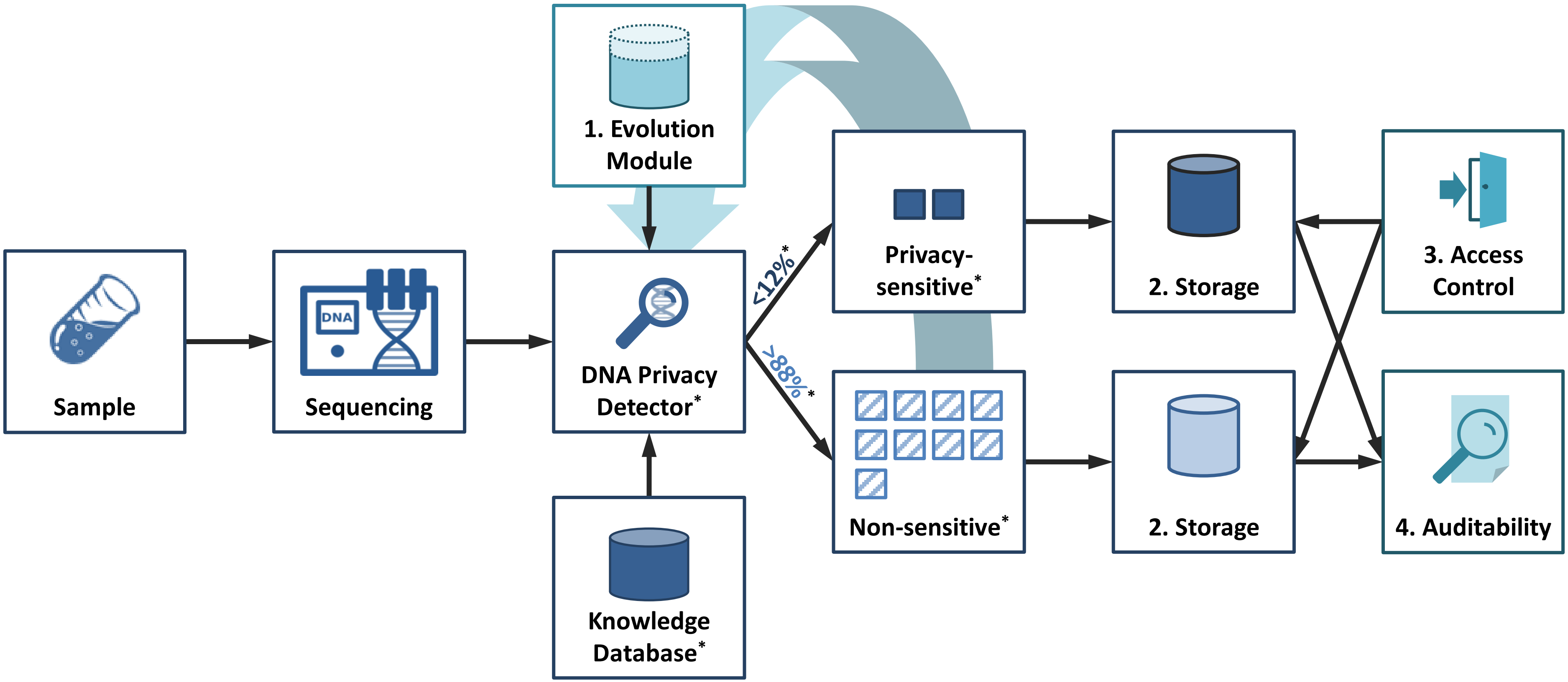
Players and interactions



1. Donors provide biological samples
2. Donors inform their sharing preferences
3. Samples are sent to sequencing
4. Sequencing data is stored in the data storage
5. Data consumer requests access and reads data
6. Auditors verify who reads what and when

2. Informed model – sharing genomes

Overview



* A High-Throughput Method to Detect Privacy-Sensitive Human Genomic Data @ WPES 2015

Evolution module

- Evolve the detection of privacy-sensitive portions
- Update when new privacy-sensitive sequences are identified
- Automatically track public databases to evolve
- Re-analyze stored data to adapt to discoveries
- Adapt access-rules
- Protect the access history for future inquires

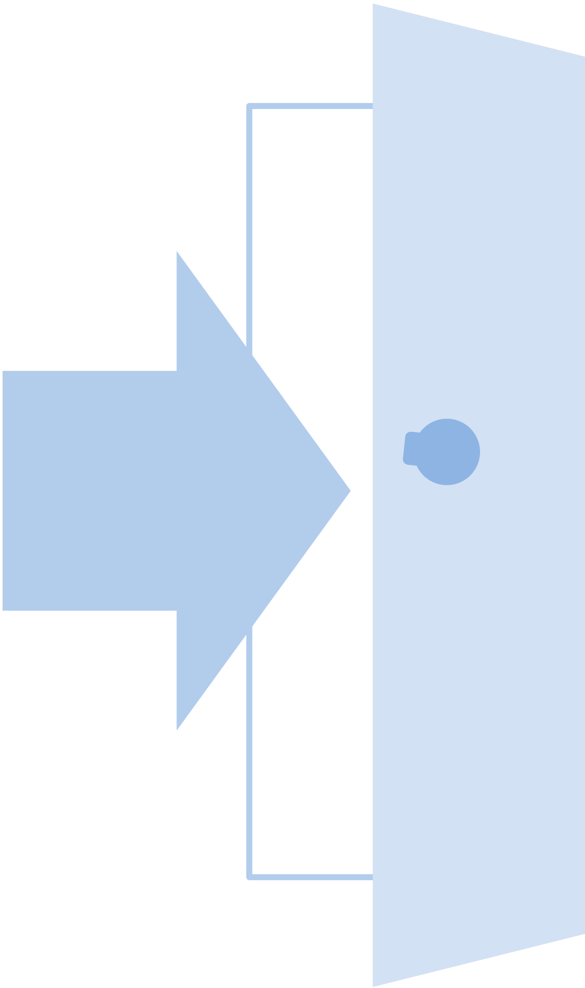
Storage

- Large amounts of data
- Use many infrastructures (i.e., from private to public clouds)
- Privacy-aware (some portions are more privacy-sensitive than others)
  - privacy-sensitive portions: appropriate levels of security and dependability
  - non-sensitive portions: affordable infrastructures
- Cost efficiency (focusing on sensitive data → reducing costs)
- Different security and dependability requirements: encryption, information dispersal, data replication, etc.
- Depend on legal constraints (e.g., outsourcing data storage may be not allowed in some countries)

2. Informed model – sharing genomes

Access Control

- Differential access based on roles and analyses
- Verify and permit researchers to access data
- Complements the evolution module
  (adapting to evolution)
- Using multiple authentication factors
- Crypto complements access control
  (intrusion does not necessarily obtain data in clear)



2. Informed model – sharing genomes

Auditability

- Audit to detect, analyze, and sanction misuses
- Who accessed what data in a chronological order?
- Auditors need only metadata, logs, and rules
- Complements the evolution module
  (who has read a previous version of a dataset)
- Supplemented by accountability (ensure all actors and actions performed are persistently recorded as evidence)
- Requires an indelible tamper-proof track of data accesses

Outline

Context

1. Analogy

2. Informed model

Final remarks

Conclusion

1. Analogy

Presented an analogy between the privacy aspects of sharing photos and sharing genomes, which clarifies the privacy risks in the latter to the general public.

2. Informed model

Illustrated an alternative informed model to share genomic data according to the privacy-sensitivity level of each portion.

Final remarks

Open questions

- **How to avoid an unreasonable** increased **management effort**?
- **How can public clouds** be securely **used** in this model?
- **Which additional** type of genomic **data** is **privacy-sensitive**?
- How **methods** associating **genomic** information and **photographic** records may impact the model?

International Conference PACBB

Thank you.

Vinicius V. Cogo[1], Alysson Bessani[1], Francisco M. Couto[1]
Margarida Gama-Carvalho[2]
Maria Fernandes[3], Paulo Esteves-Verissimo[3]

[1]LASIGE, University of Lisbon
[2]BioSys, University of Lisbon
[3]SnT, University of Luxembourg

U LISBOA | UNIVERSIDADE DE LISBOA

uni.lu UNIVERSITÉ DU LUXEMBOURG