Bocconi

A (BRIEF) OVERVIEW ON PRIVACY ENHANCING TECHNOLOGIES (PETs)

Emmanuela Orsini

Università Bocconi
MILANO

The current paradox of data economy

- The exponential increase in the processing of personal data has created a wide array of unprecedented possibilities to gain useful insights via artificial intelligence and machine learning; at the same time, these **developments expose individuals to new privacy threats.**

- Privacy **vs** usability

Privacy enhancing technologies

PETs want to bridge the gap of data protection on one side and the value of data on the other, also called the **privacy-utility tradeoff**.

Legislator and public authorities on PETs

The US Senate in "Promoting Digital Privacy Technologies Act"

- Proposes the definition of PETs as "any software solution, technical processes, or other technological means of enhancing the privacy and confidentiality of an individual's personal data in data or sets of data".

- The Act is planning to provide support for research, broader deployment and standardization

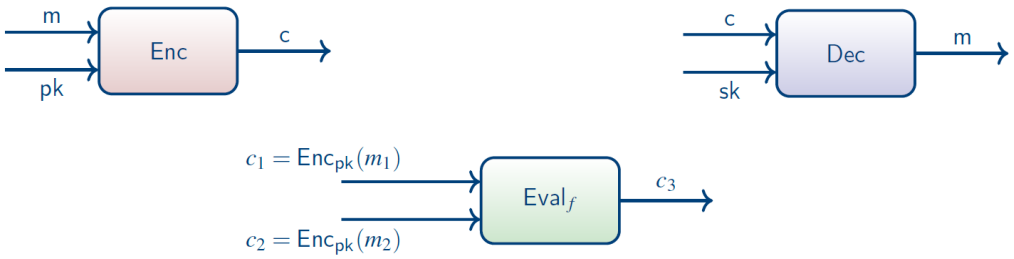
PETs categories

Types of PETs	Key technologies	Current and potential applications*	Challenges and limitations
Data obfuscation tools	Anonymisation / Pseudonymisation	Secure storage	- Ensuring that information does not leak (risk of re-identification) - Amplified bias in particular for synthetic data - Insufficient skills and competences
	Synthetic data	Privacy-preserving machine learning	
	Differential privacy	Expanding research opportunities	
	Zero-knowledge proofs	Verifying information without requiring disclosure (e.g. age verification)	- Applications are still in their early stages
Encrypted data processing tools	Homomorphic encryption	Computing on encrypted data within the same organisation Computing on private data that is too sensitive to disclose Contact tracing / discovery	- Data cleaning challenges - Ensuring that information does not leak - Higher computation costs
	Multi-party computation (including orivate set intersection)		
	Trusted execution environments	Computing using models that need to remain private	- Higher computation costs - Digital security challenges
Federated and distributed analytics	Federated learning	Privacy-preserving machine learning	- Reliable connectivity needed - Information on data models need to be made available to data processor
	Distributed analytics		
Data accountability tools	Accountable systems	Setting and enforcing rules regarding when data can be accessed Immutable tracking of data access by data controllers	- Narrow use cases and lack stand-alone applications - Configuration complexity - Privacy and data protection compliance risks where distributed ledger technologies are used - Digital security challenges - Not considered as PETs in the strict sense
	Threshold secret sharing		
	Personal data stores / Personal Information Management Systems	Providing data subjects control over their own data	

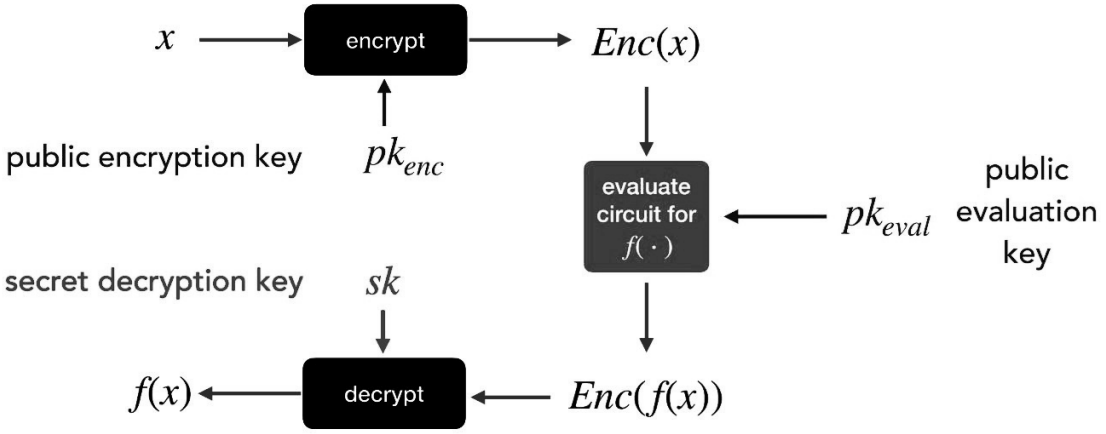
PETs categories

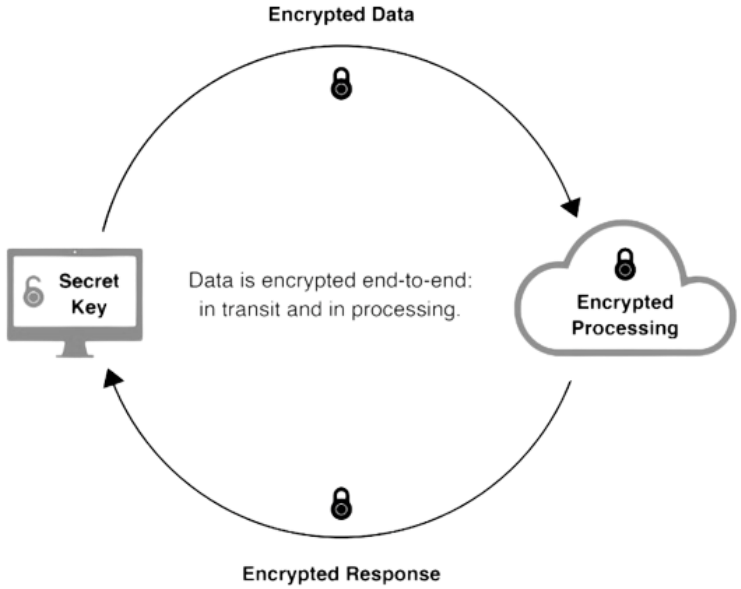
Types of PETs	Key technologies	Current and potential applications*	Challenges and limitations
Data obfuscation tools	Anonymisation / Pseudonymisation	Secure storage	- Ensuring that information does not leak (risk of re-identification) - Amplified bias in particular for synthetic data - Insufficient skills and competences
	Synthetic data	Privacy-preserving machine learning	
	Differential privacy	Expanding research opportunities	
	Zero-knowledge proofs	Verifying information without requiring disclosure (e.g. age verification)	- Applications are still in their early stages
Encrypted data processing tools	Homomorphic encryption	Computing on encrypted data within the same organisation Computing on private data that is too sensitive to disclose Contact tracing / discovery	- Data cleaning challenges - Ensuring that information does not leak - Higher computation costs
	Multi-party computation (including orivate set intersection)		
	Trusted execution environments	Computing using models that need to remain private	- Higher computation costs - Digital security challenges
Federated and distributed analytics	Federated learning	Privacy-preserving machine learning	- Reliable connectivity needed - Information on data models need to be made available to data processor
	Distributed analytics		
Data accountability tools	Accountable systems	Setting and enforcing rules regarding when data can be accessed Immutable tracking of data access by data controllers	- Narrow use cases and lack stand-alone applications - Configuration complexity - Privacy and data protection compliance risks where distributed ledger technologies are used - Digital security challenges - Not considered as PETs in the strict sense
	Threshold secret sharing		
	Personal data stores / Personal Information Management Systems	Providing data subjects control over their own data	

(Fully) homomorphic encryption (FHE)

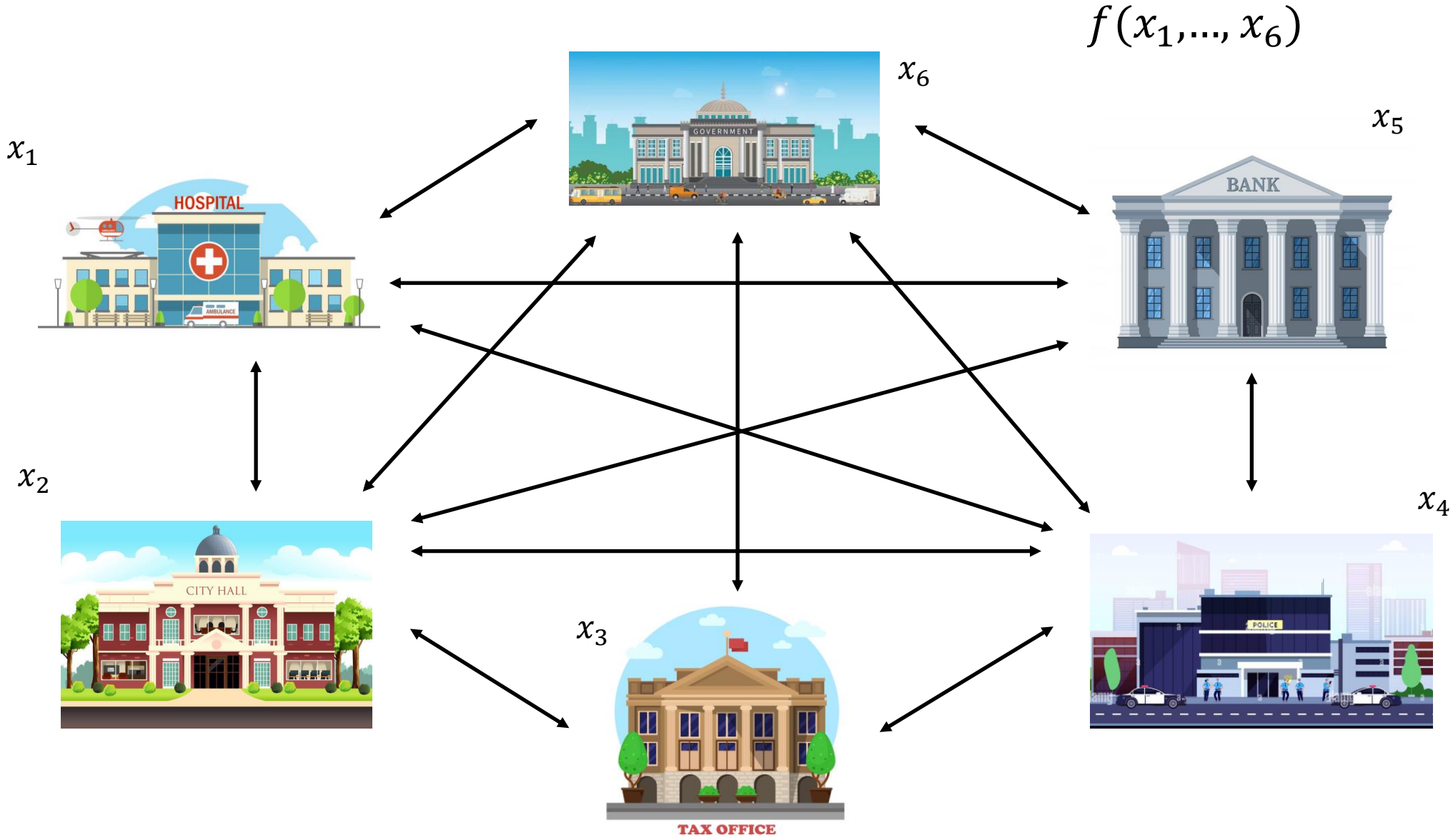


$$\mathsf{Dec}_{\mathsf{sk}}\left(\mathsf{Eval}_f(\mathsf{ek}, c_1, c_2)\right) = f(m_1, m_2)$$

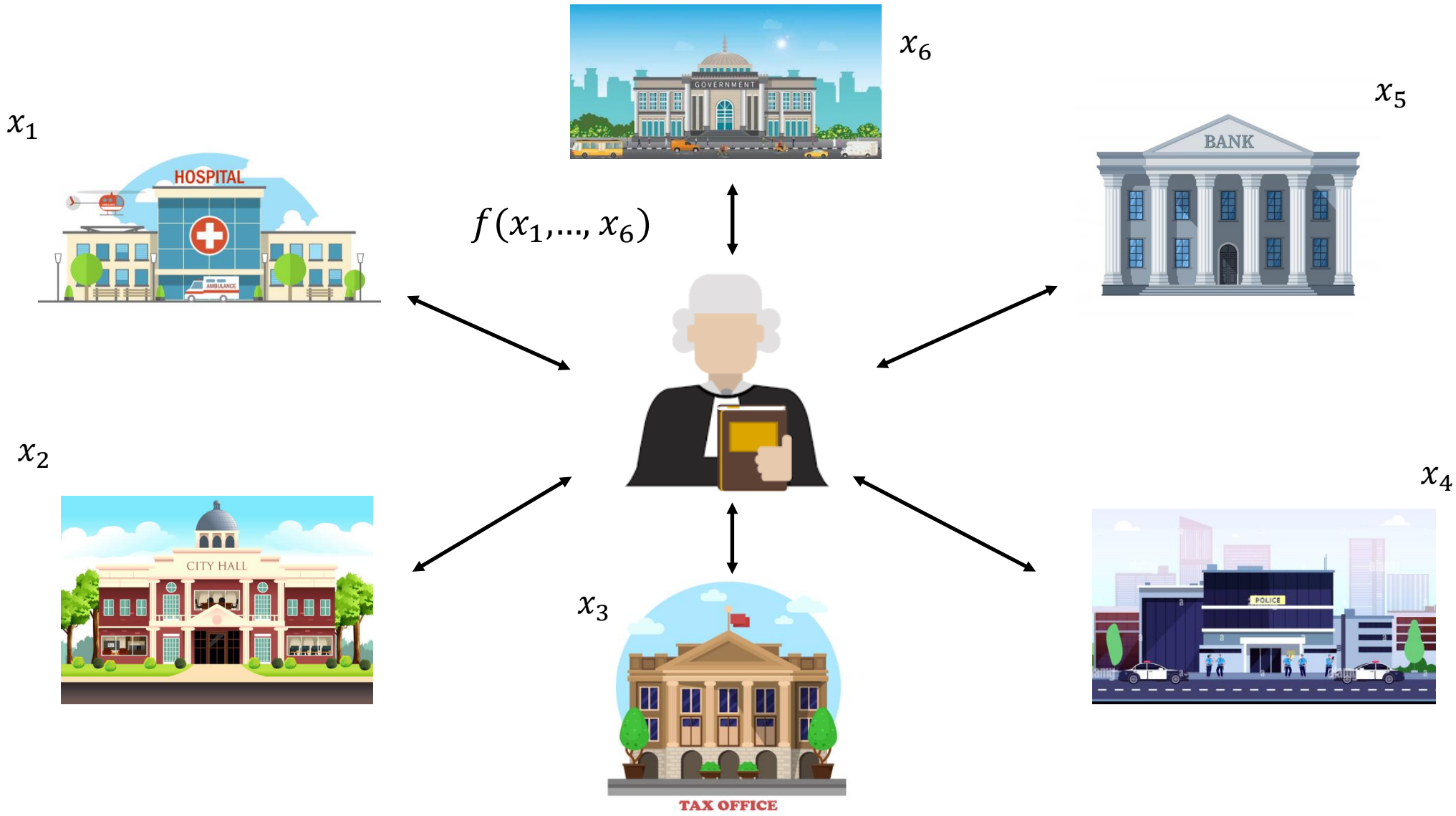




Secure multiparty computation (MPC)

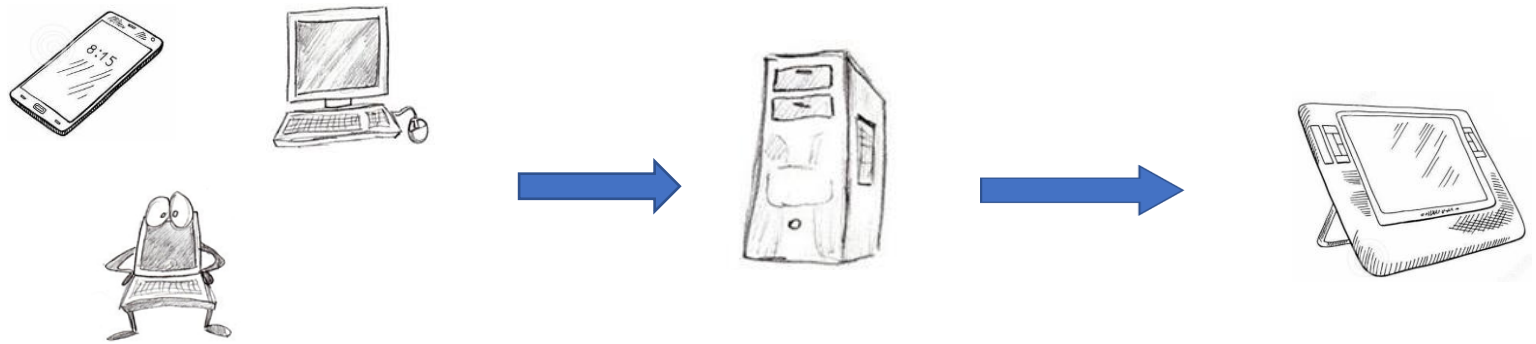


Secure multiparty computation

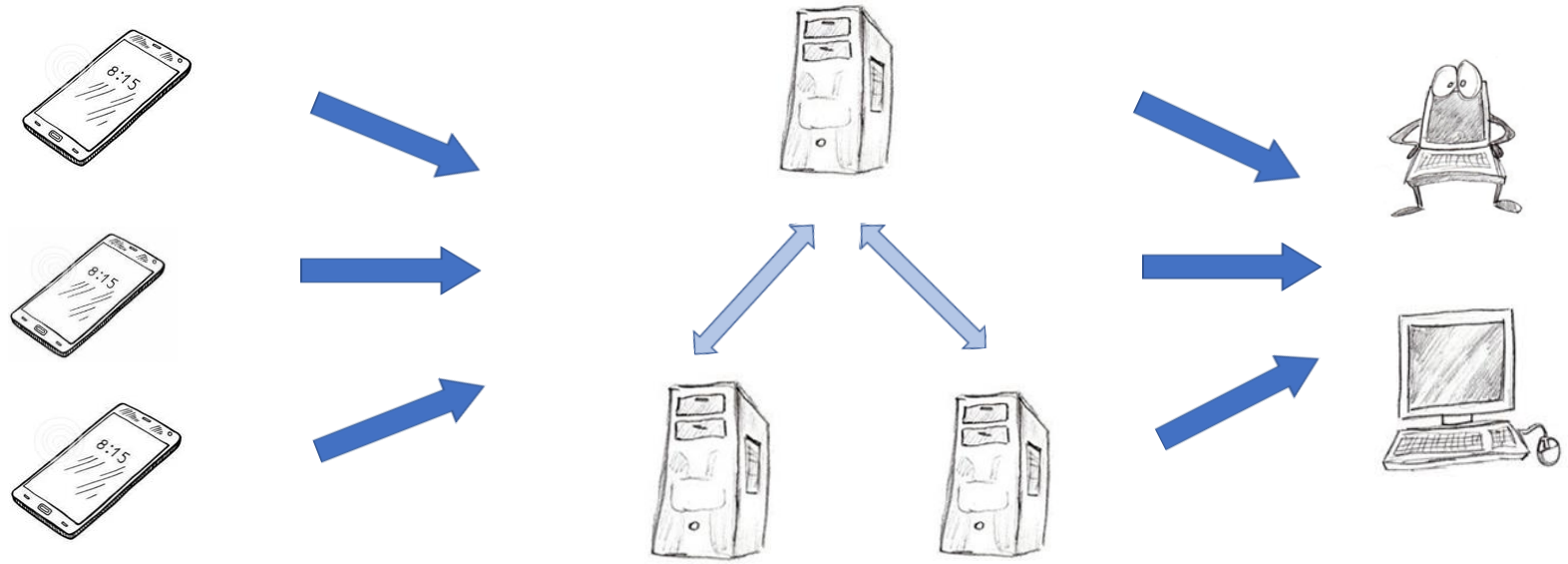


Secure multiparty computation vs FHE

FHE



MPC



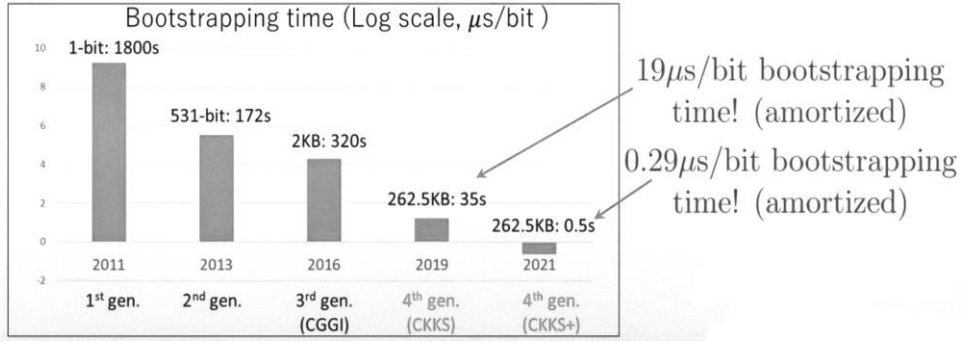
Is this even possible?

FHE

"I don't think we'll see anyone using Gentry's solution in our lifetimes."

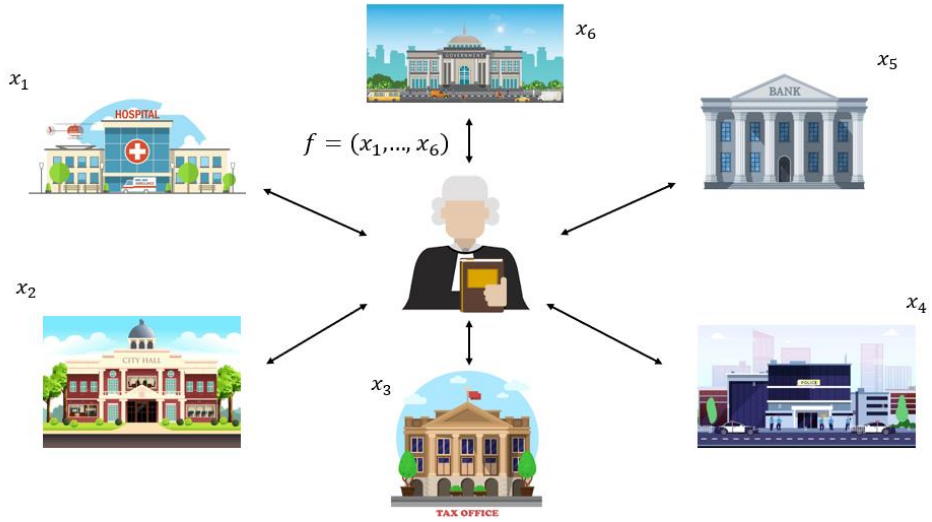
- HE is getting faster 8 times every year

  e.g. Bootstrapping time: the most time-consuming operation in HE



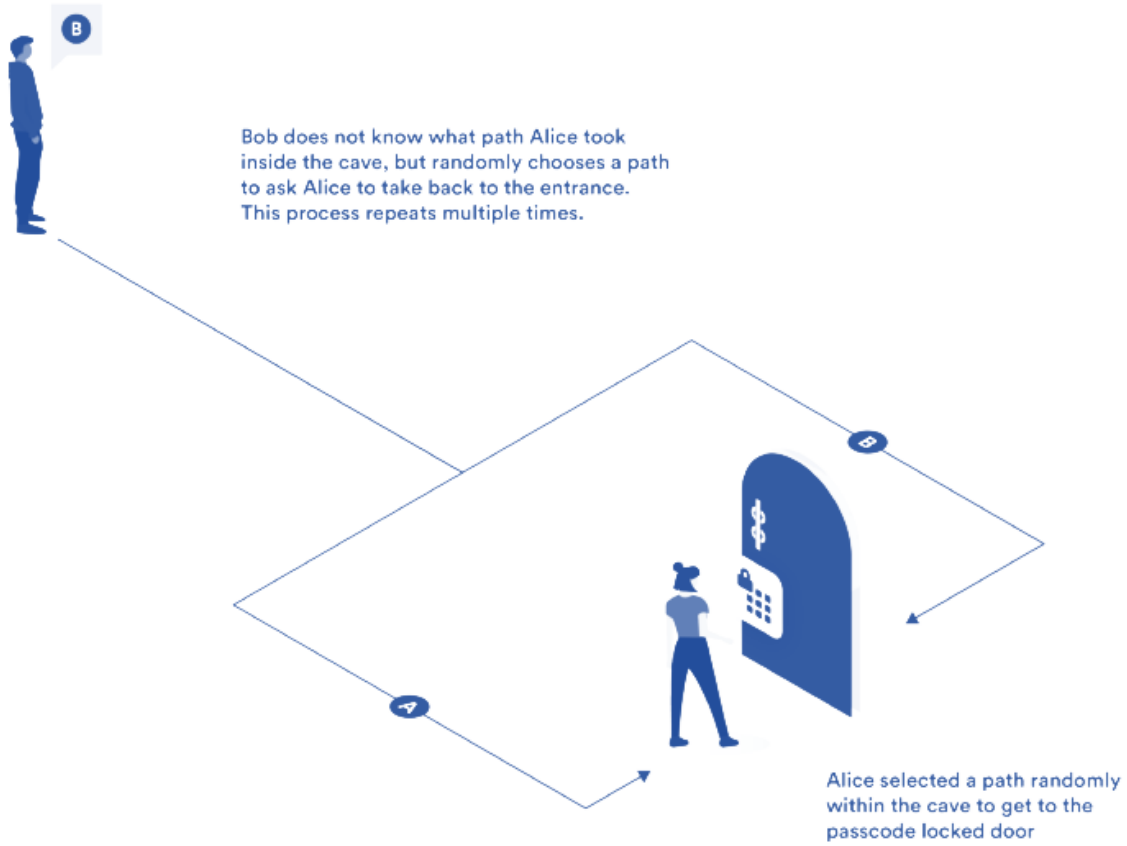
- Still slow in computation
- Relatively cheap in communication
- Only possible (currently) for simple functions

MPC



- MPC is a very mature technology
- Several general purpose protocols
- It allows collaborative computing, hence it requires communication between parties
- Generally cheap in computation

Zero-knowledge proofs



- ZKPs enable a party to demonstrate a statement to other parties without revealing any information beyond the proof.
- ZK Proofs facilitate the development of an entirely new range of applications, e. g. blockchain, “proof of humanity”, national security purpose

Zero-knowledge proofs – The SIEVE project

(Securing information for encrypted verification and evaluation)

Defense Advanced Research Projects Agency: ``Recent research has substantially increased the efficiency of ZK proofs, enabling real-world use, primarily by cryptocurrencies. While useful for cryptocurrencies, the ZK proofs created are specialized for **this task and do not necessarily scale for transactions that are more complex.** For highly complex proof statements like those that the Department of Defense (DoD) may wish to employ, novel and more efficient approaches are needed.''

1. Prove whether or not vulnerabilities exist in software without revealing how they are caused;
2. Assess properties of cyberspace operations capabilities without accessing source code;
3. Assure anti-tamper cyber-security techniques without revealing code details.

ZK proofs and post-quantum signature schemes

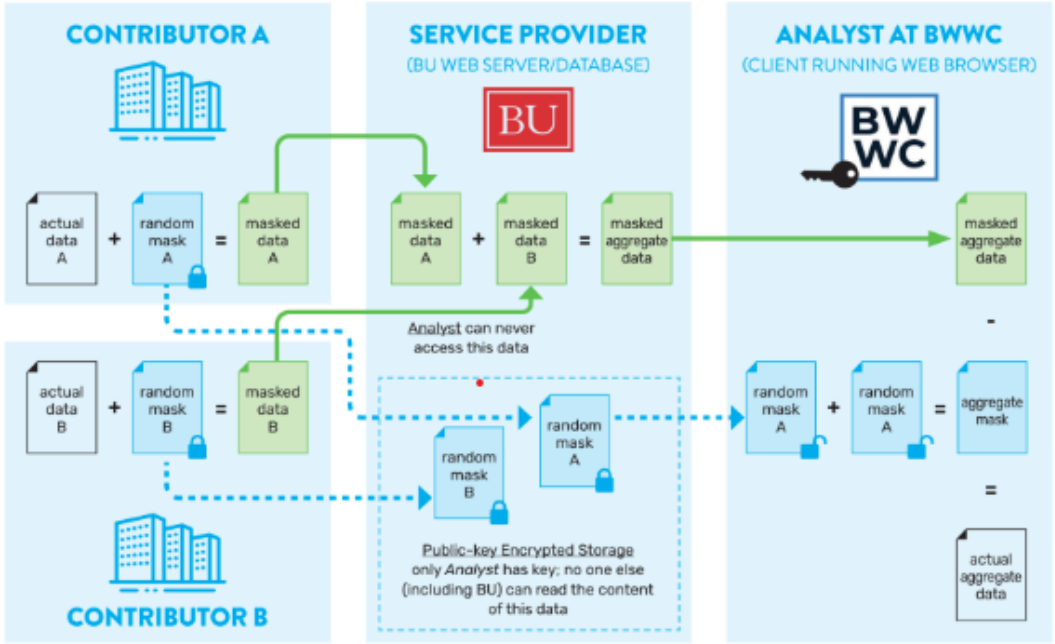
F*AEST*: new post-quantum signature scheme

- Only based on symmetric primitives (AES-SHAKE)
- Small keys
- Faster in signing and shorter than SPHINCS+
- Based on MPC tools

It will be submitted to the next NIST PQC standardization call on the 1st June.

Case study report - UN Committee on Big Data and Data Science for Official Statistics

Boston Women's Workforce Council: Measuring salary disparity using secure multi-party computation

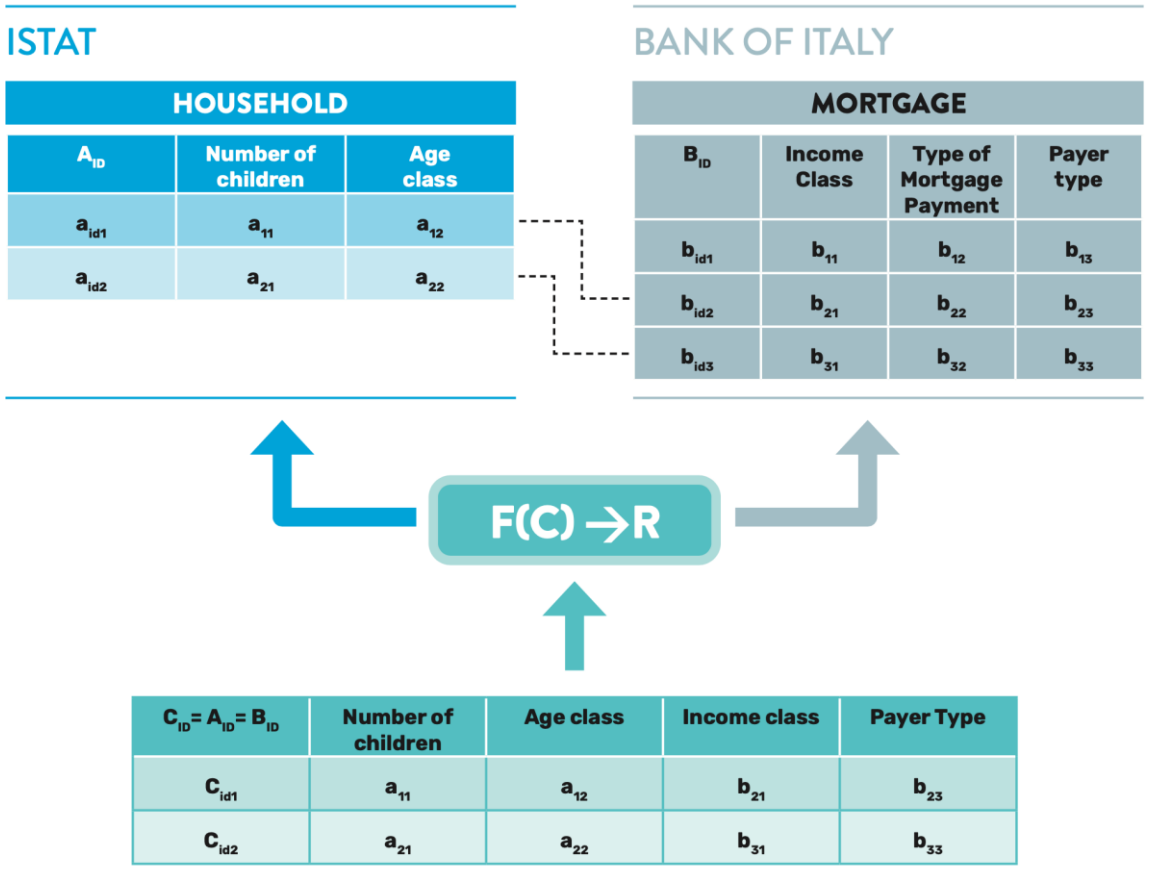


Purpose	To measure the gender and racial wage gaps throughout the greater Boston area every 1-2 years
Datasets	Real demographic and payroll data from companies and non-profit organisations, large and small, throughout the greater Boston area
PETs used	Secure Multi Party Computation
Application	Secure vector addition
Details of computation	Organisations contribute a spreadsheet containing more than 600 cells of data. The Boston Women's Workforce Council receives the summation of each cell across all participating organisations.
Parties and trust relationship	More than 100 participating organisations act as input parties; the Boston Women's Workforce Council serves in a compute and output party role; Boston University serves as a compute party. Participants trust BU and the BWWC to behave semi-honestly, with the ability to audit and verify code.
Implementation status	Production
Resources	Boston Women's Workforce Council reports Data submission website Open-source code repository on GitHub Publications about the PET used in this project appear at SOUPS 2019, COMPASS 2018, SecDev 2016, and the Communications of the ACM.

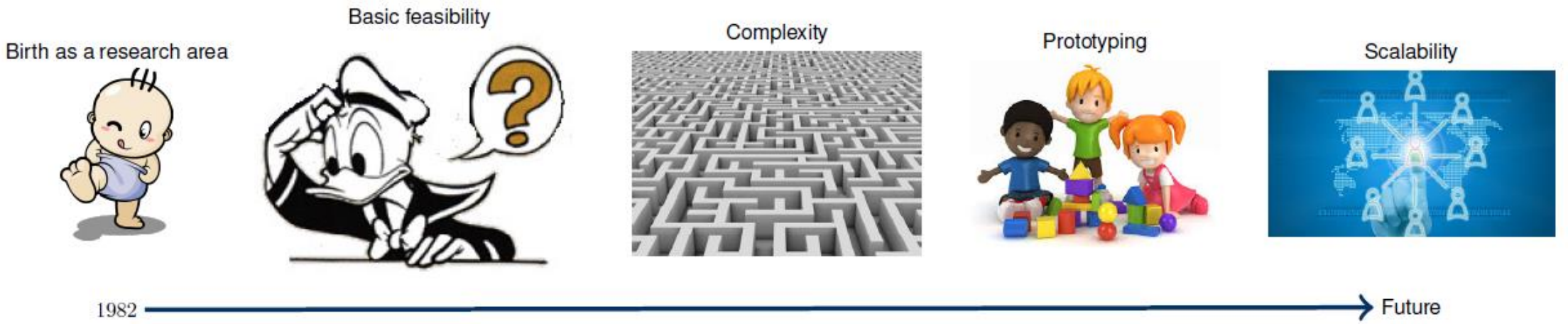
Case study report - UN Committee on Big Data and Data Science for Official Statistics

Italian National Institute of Statistics and Bank of Italy: Enriching data analysis using privacy-preserving record linkage

Purpose	To enable enriched socio-economic analysis by augmenting data held by Bank of Italy with data held by ISTAT (and vice versa).
Datasets	Socio-demographic and financial datasets (linkable via tax code common key)
PETs used	Secure Multi Party Computation
Application	Private Set Intersection with Analytics
Details of computation	ISTAT and Bank of Italy perform an Exact PSI using the shared tax code key. The intersection is encrypted and transferred to the third "linker" party. ISTAT and Bank of Italy submit queries to the linker, which can perform aggregation and counts against the data on-demand, with outputs transmitted to ISTAT and Bank of Italy.
Parties and trust relationship	Bank of Italy and ISTAT act as input, compute, and output parties; third "linker" party serves in a compute role. Organisations trust each other ("honest but curious" threat model).
Implementation status	Pilot
Resources	



Future directions



All the stages in this line are still very active!!!

PETs maturity, opportunities and challenges

- PETs **are at different stages** of development and will likely need to be part of data governance frameworks to ensure they are used properly. **Many of these tools are still in their infancy and limited to specific data processing use cases**.

- Given their innovative nature and high potential, **PETs warrant a comprehensive re-evaluation of the application of regulations on data collection and processing**

- **Policy makers will increasingly need to consider how the use of PETs may impact regulatory assessments under national privacy and data protection frameworks**, taking into account the contribution of PETs to privacy protective outcomes.

- As PETs mature, there will be an increasing need for awareness raising and training to better design, build, implement, use and audit these new technologies.

- Stronger cross-border and cross-sectoral regulatory co-operation will be needed to better consider technological developments on PETs for privacy and data protection.

DOMANDE ?