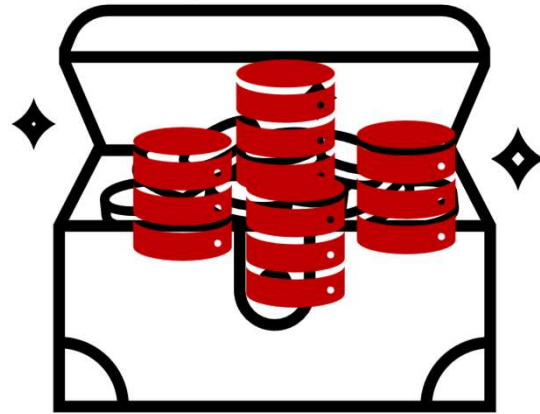


Balancing privacy and utility: anonymisation and pseudonymisation

By Noëmi Debacker and Ine Steyls
National Bank of Belgium

A lot of interesting research data at the National Bank of Belgium (NBB)

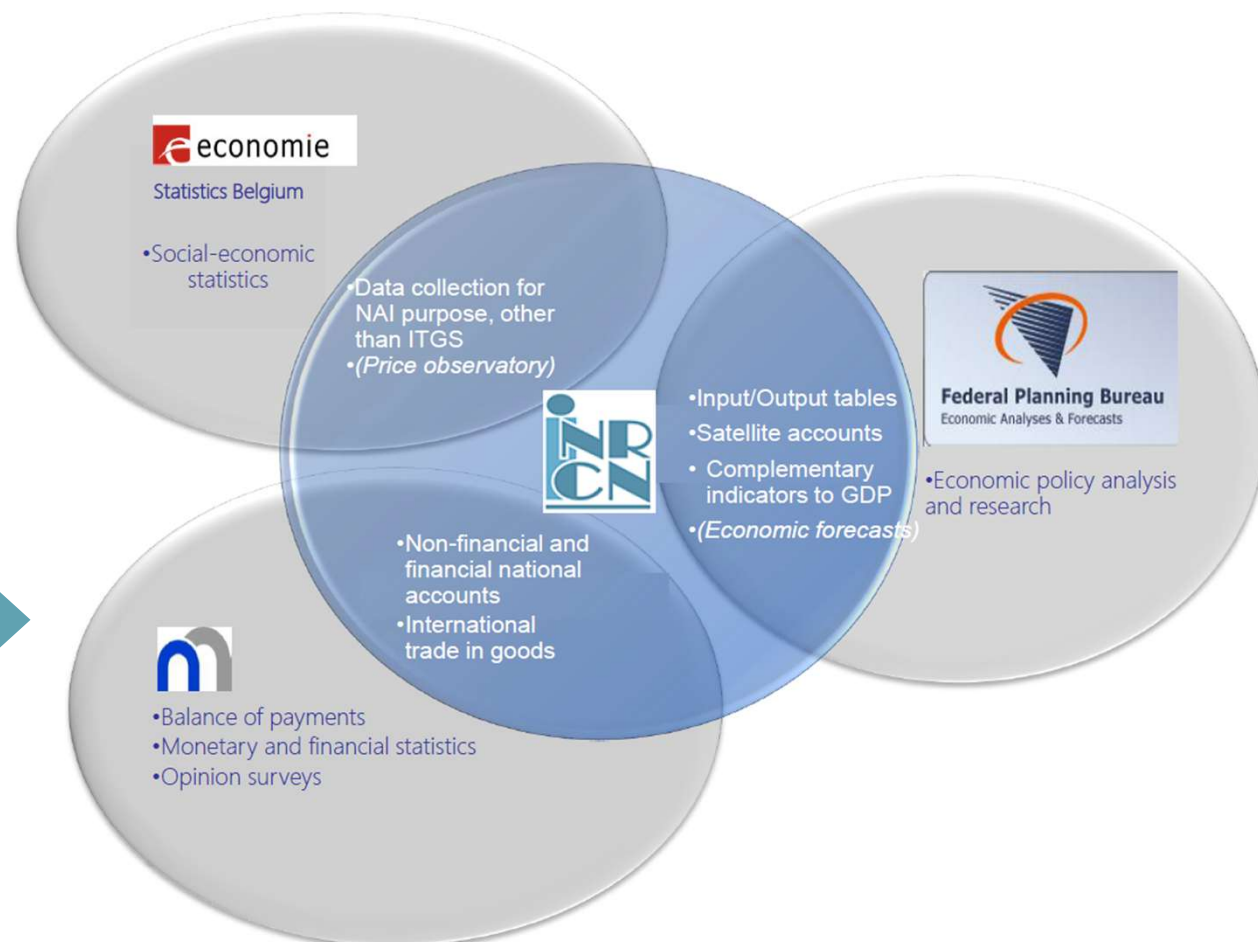


- collected by the NBB, sourced from external providers, bought, ...
- within the framework of legal obligations; daily operational tasks and duties; statistical purposes, research purposes, ...
- public (business register and companies' financial statements), confidential (major part) and personal data

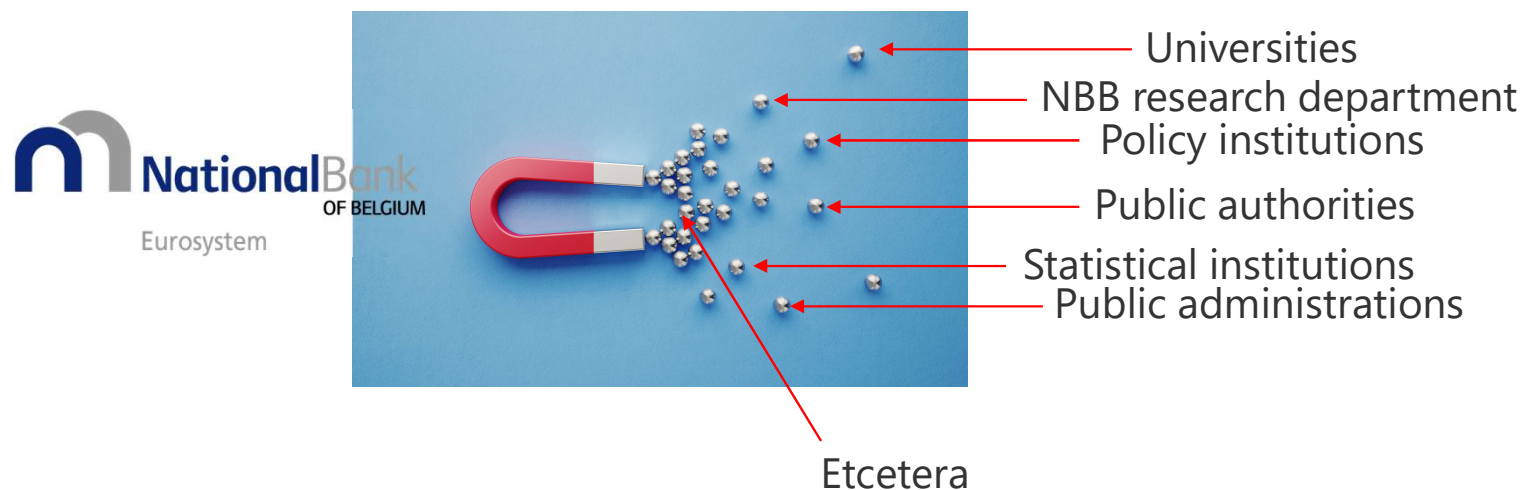
... including statistical data

- Belgium has a complex statistical landscape composed by different organisations being involved, among which the NBB
- Financial statistics = ECB statistics to support Euro Area.
- The national accounts are produced by the National Accounts Institute (**INR-ICN**), a central coordinating body composed by the NBB, the Federal planning bureau and Statbel.

Specific
to
Belgium



Processing data for purposes beyond the initial purpose



- The possibility to share data depends on the type of data, the legal framework – among which GDPR & statistical law –, who requests the data, legitimate purpose, conditions set by the data provider, where the data will be processed, ...
- Focus of this presentation: requests for data for research or policy support
- For a lot of data (but not all (!), e.g. prudential data) confidential data can be used after removal of direct identifiers, i.e. pseudonymous data

Anonymisation vs Pseudonymisation:

Core Concepts we use at the NBB and in this slide

- Anonymous data

All identifying elements are irreversibly removed → re-identification of an individual entity is not possible by any means existing today.

- Pseudonymous data

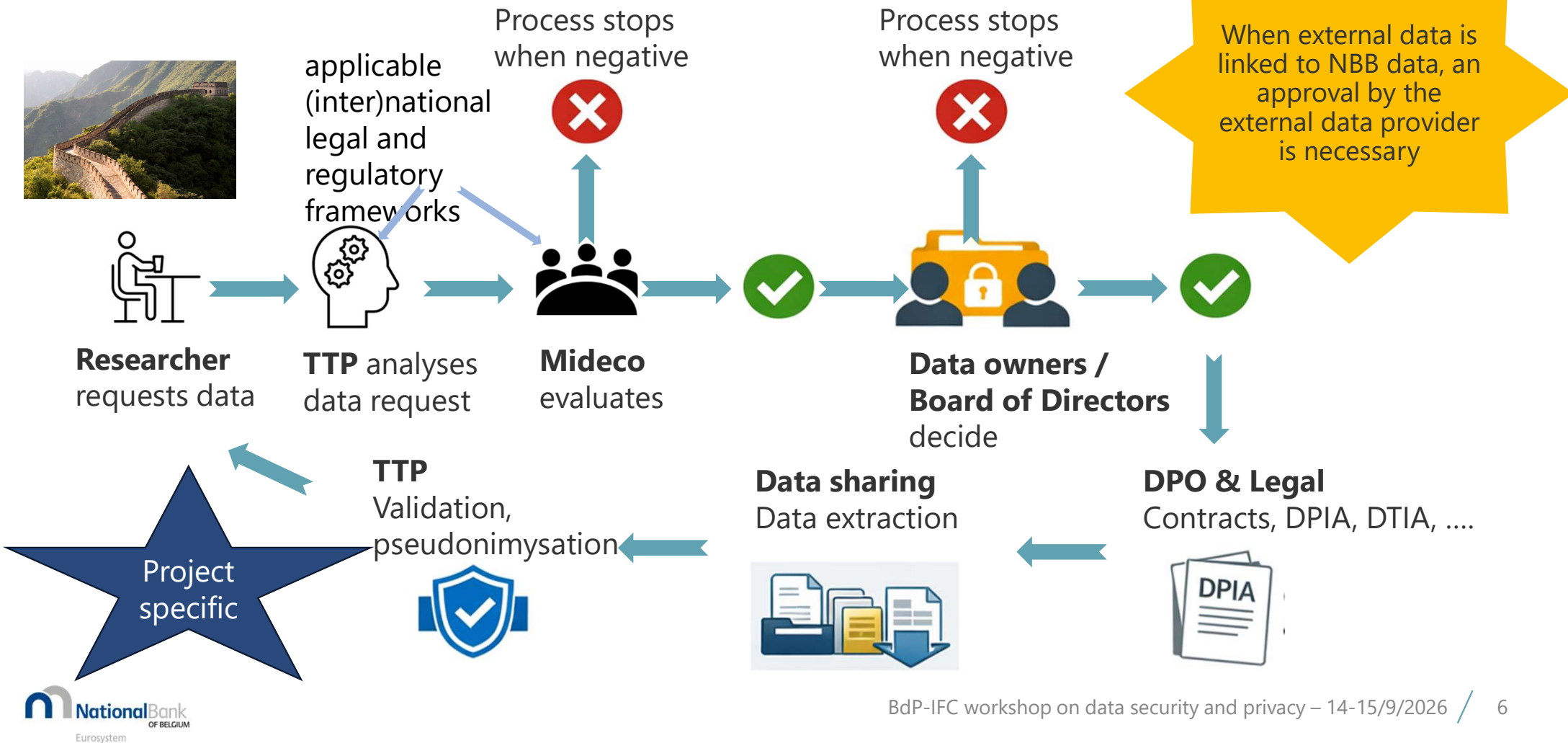
- Direct identifiers are replaced with pseudonyms while maintaining a one-to-one relationship between the original identity and the pseudonym → indirect identification is possible.

- Data in which pseudonyms are removed but still allow indirect identification by combining information in the data set are pseudonymous data and are **not** anonymous data!

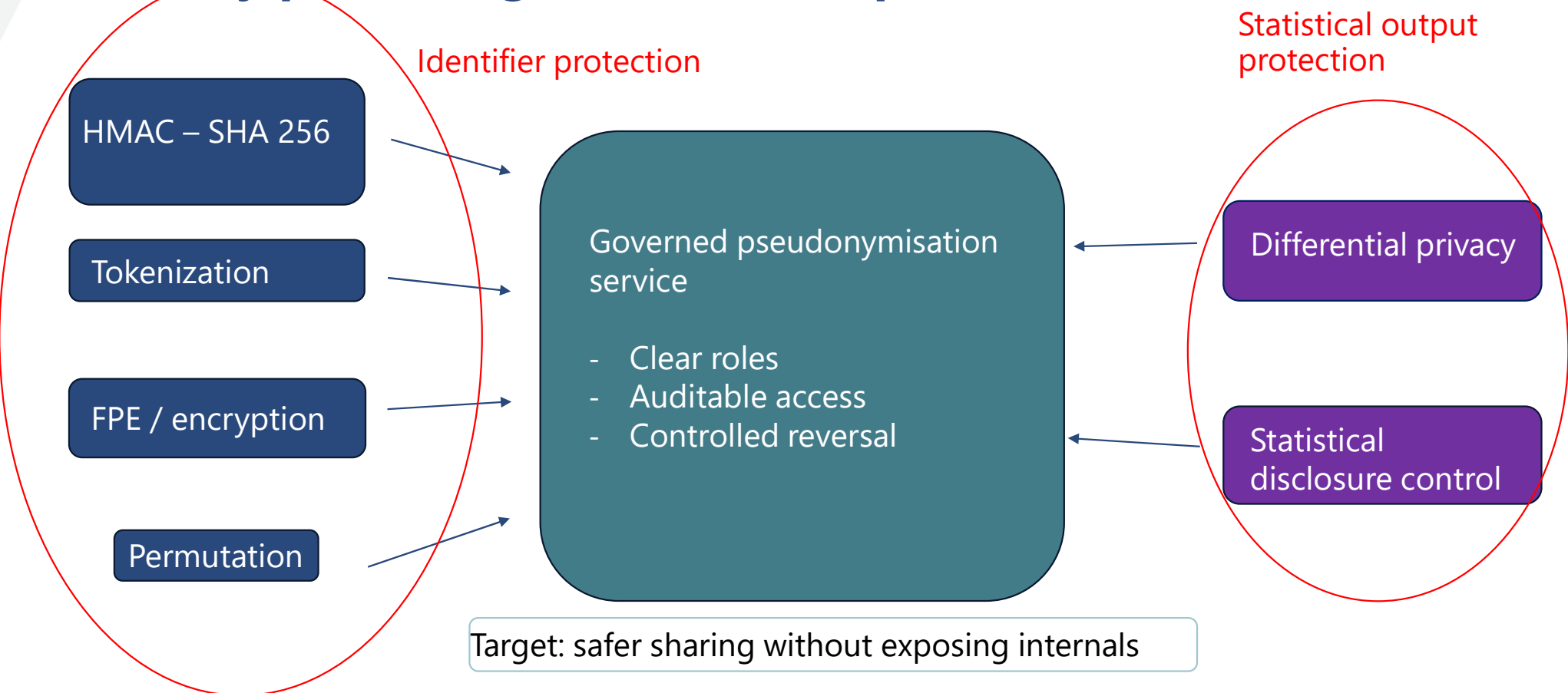
- Is pseudonymisation alone enough



Data Access Framework at NBB



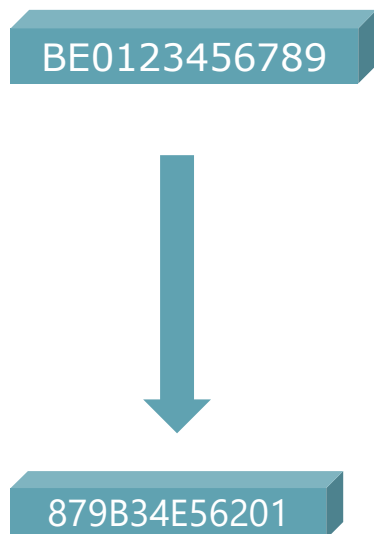
Privacy preserving architecture options



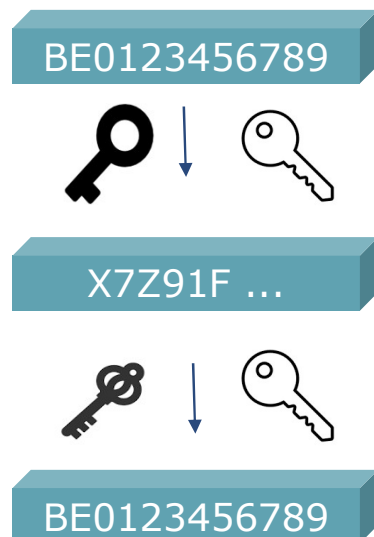
Techniques protect the data, Governance protects the process

Techniques for protecting identifiers

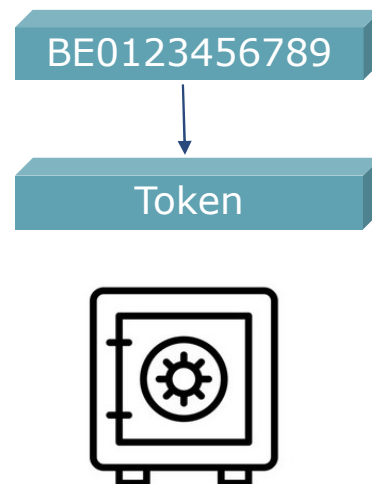
Permutation



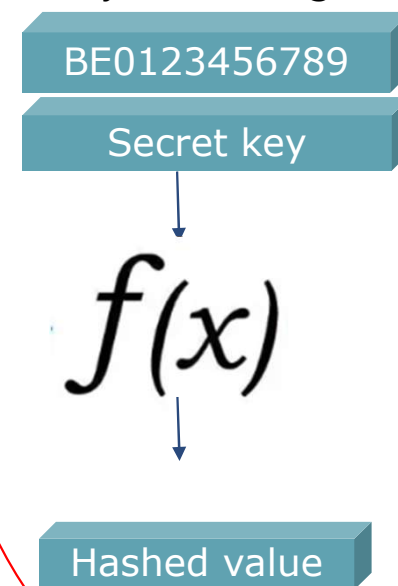
Encryption



Tokenization



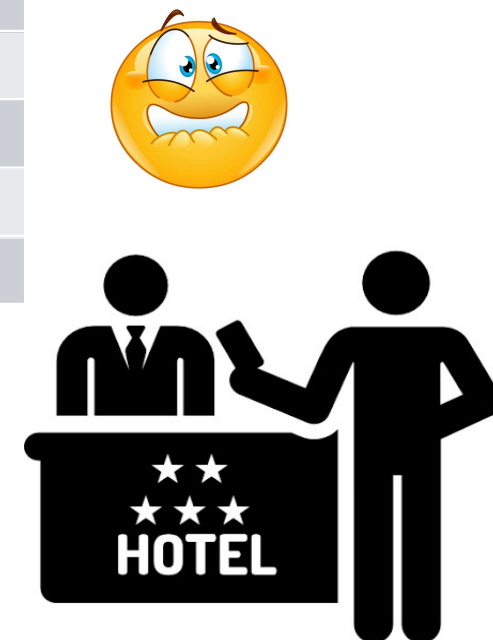
Keyed Hashing



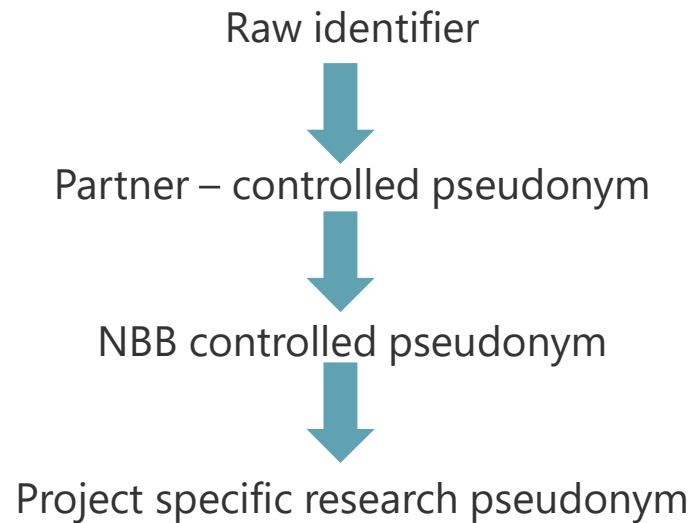
Toy Hashing – example with easy function

Name	ASCII sum	Mod 6	Room
TRUMP	408	0	0
PUTIN	400	4	4
ZELENSKY	637	1	1
MACRON	448	4	4
RUTTE	404	2	2

Real cryptographic constructions do not eliminate the mathematical possibility of collisions, but they provide a vastly larger and more uniformly distributed output space



R-package: "nbbpseudonymize"



Researchers can link records consistently without receiving the original company identifier

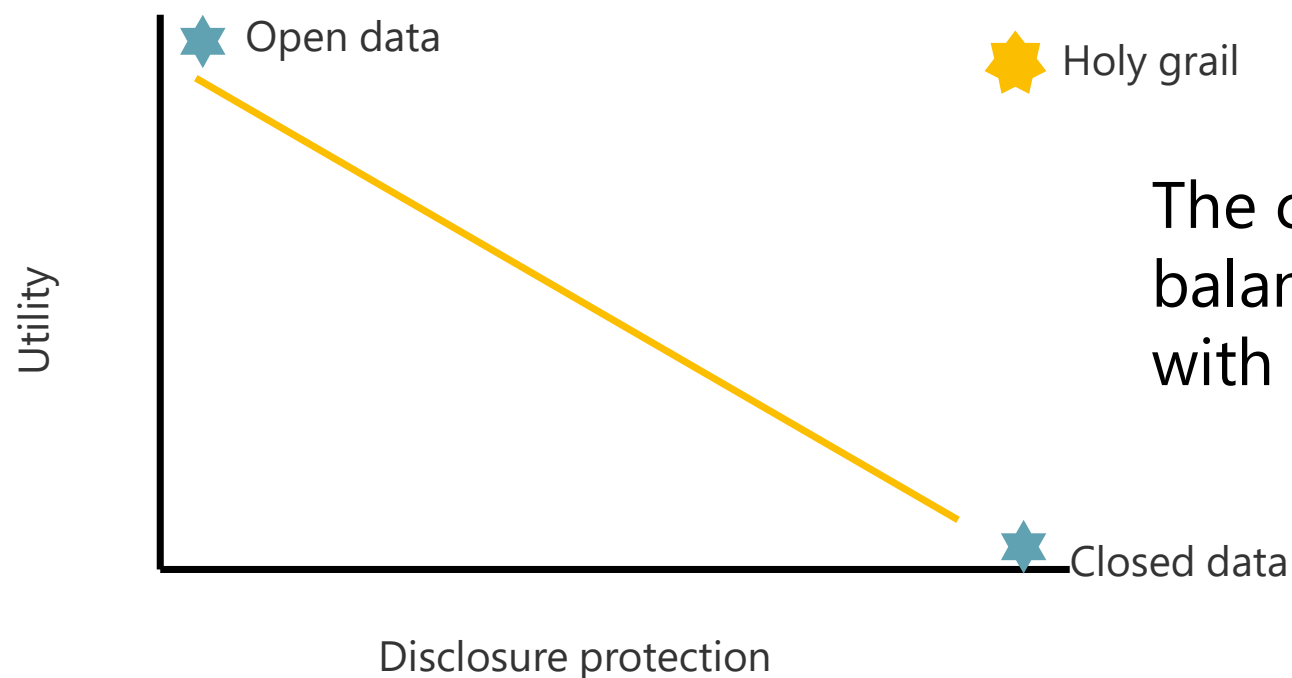
Challenges: Indirect Identification and Public Data

- Public data in Belgium:
 - Belgian Crossroads Bank for Enterprises (official business register)
 - Central Balance Sheet Office (CBSO) data (companies' financial statements)
- Increase in the risk of indirect identification
 - How to introduce noise to the data without impacting the usability of the data too much (Privacy–Utility Trade-off)?
 - Should we think of systems with no direct access to the data as such, e.g. remote execution or remote access without possibility of viewing the microdata?
 - Feasibility?

Beyond Techniques: Governance and Responsible Data Use

- Data protection is not only a technical challenge: **human factors matter**
- Continuous awareness is required from everyone involved in handling confidential data to:
- protect confidentiality throughout the data lifecycle by:
 - understanding that data may remain confidential in different forms, including detailed and aggregated data
 - recognising confidentiality obligations
 - using data only for legitimate purposes
 - following established procedures and controls
 - Goal: prevent misuse and unauthorised disclosure, whether intentional or accidental

The Privacy–Utility Trade-off



The challenge:
balancing legal compliance
with research freedom