



PROTECTING DATA AND ITS VALUE: PRESERVING ANALYTICAL UTILITY

MÁRIO LOURENÇO
FRANCISCO FONSECA

WORKSHOP ON "DATA SECURITY AND PRIVACY"
LISBON, SEPTEMBER 2026



BANCO DE
PORTUGAL
EUROSISTEMA



WHY IT MATTERS

01



WHY IT MATTERS

A GROWING CONCERN

Central Banks have access to a **growing number of datasets** with **microdata** concerning **companies** and **natural persons**...

Central Credit Registers
Bank Accounts Database

... benefiting from **directly reported data** as well as **administrative databases**...

Business Registers, Tax Authorities Data
(Income, Social Security, etc.)

... promoting **increased data integration** and **data sharing**

Integrated Data Management Programmes
Data Warehouses
Quick and easy access to all data

... while also being able to ensure **regulatory compliance**...

GDPR, AI Act

... and **maintain public trust**

Avoid **wrongful disclosure of data**
(*even when there is pressure to be agile*)
Protect the identity of all individuals
whose **data we are entrusted with**



- How do we balance the need to **protect data** while keeping it available to be easily **used for different purposes**?

**Protection should be proportional
to purpose, risk and
expected analytical use**

FROM PSEUDONYMIZATION TO RISK-BASED DATA PROTECTION

02



FROM PSEUDONYMIZATION TO RISK-BASED DATA PROTECTION

GENERAL PRINCIPLES

4th IFC-BdI Workshop on "DATA SCIENCE IN CENTRAL BANKING", ROME, February 2025

ANONYMIZATION PRINCIPLES MAIN GOALS



Define

Ensure that core concepts – **personal data**, **anonymization**, **pseudonymization** – have a **common definition** at the institutional level



Prototype

Idealize and implement a working prototype for an **internal pseudonymization algorithm**



Idealize

Consider the **key issues** that must be tackled when defining a data privacy model for the **department**, and eventually for the **bank**



Implement

Objectively define the steps that must be taken **to implement the pseudonymization model** and establish the corresponding **governance model**

DATA ANONYMIZATION PRINCIPLES AT BANCO DE PORTUGAL

FEBRUARY 2025 9

DEFINING THE PATH TOWARDS IMPLEMENTATION PROPOSED MODEL – MAIN TOPICS

1

Critical databases and variables

Evaluate **all databases** where the unit is an individual, and define a list of quasi-identifiers that pose a significant risk of identification

2

Define an access policy

Proposed policy comprised of **3 types of profile**, and focused on **democratizing the access to all pseudonymized data**

3

Pseudonymization algorithm

Define and implement an **internal pseudonymization algorithm** that can be applied to **all relevant databases** regardless of the types of identifier/quasi-identifiers

4

Data flow

Implement a **data flow** policy that **must be followed** when receiving, using, and sharing **any data where the unit is an individual**

DATA ANONYMIZATION PRINCIPLES AT BANCO DE PORTUGAL

FEBRUARY 2025 14

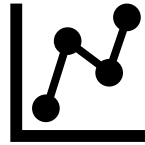
In 2025, we presented a common backbone for an anonymization model.

In 2026, the challenge is to make that backbone risk-aware, use-specific and utility-preserving.



BALANCING PRIVACY AND UTILITY

DATA PSEUDONYMIZATION



The **analytical value** of microdata lies mostly in the **detection of patterns** and **not in the individual data itself**

Data **pseudonymization** **preserves analytical value** while **mitigating identification risk**

Different **pseudonymization techniques** can be applied

Replacement of direct identifiers with pseudonyms

Generalization of variables like age and location

Suppression of high-risk variables (zip codes, birth dates)



FROM PSEUDONYMIZATION TO RISK-BASED DATA PROTECTION

DATA PSEUDONYMIZATION

Key features need to be accounted for



Pseudonyms
must be consistent
to ensure links between
databases and across time

Sufficiently **agile and flexible** to
be applicable to all relevant
datasets

Risk must be assessed
individually and collectively



But is this enough?

Is a “**one-size-fits-all**” solution adequate?
Or do we need a “**case-by-case**” approach?

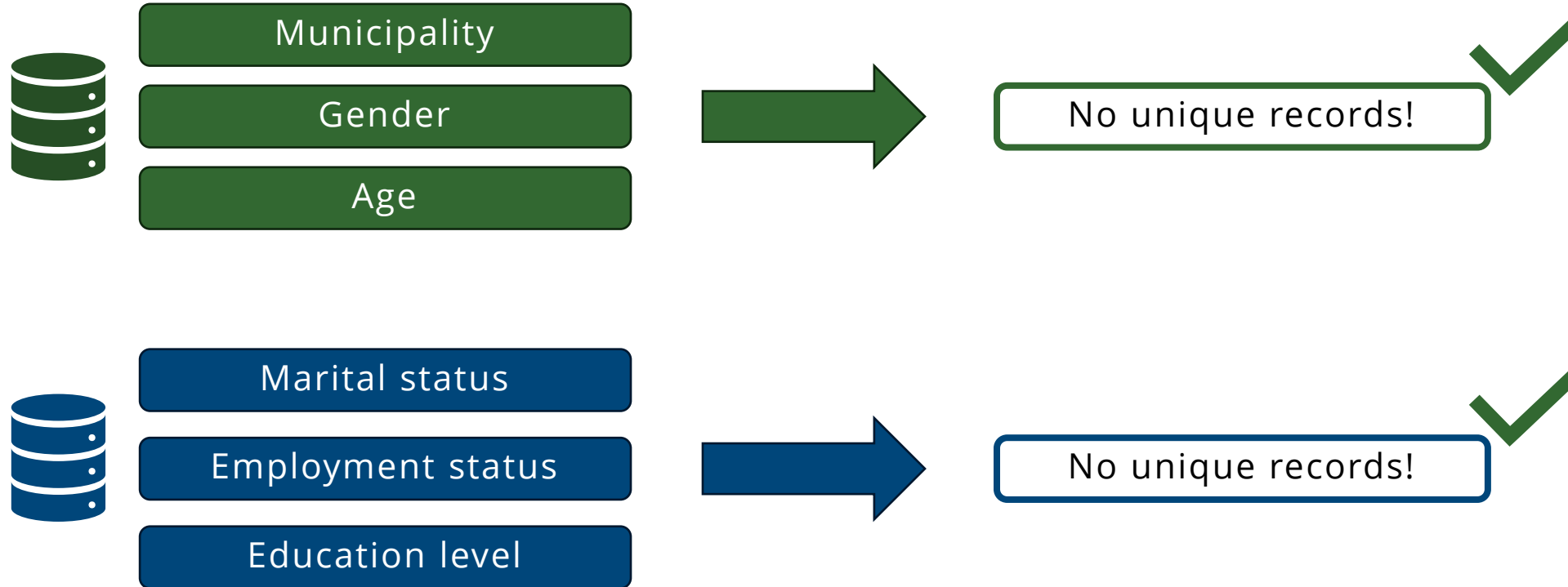
THE REIDENTIFICATION RISK ISSUE

03



THE REIDENTIFICATION RISK ISSUE

A SIMPLE EXAMPLE



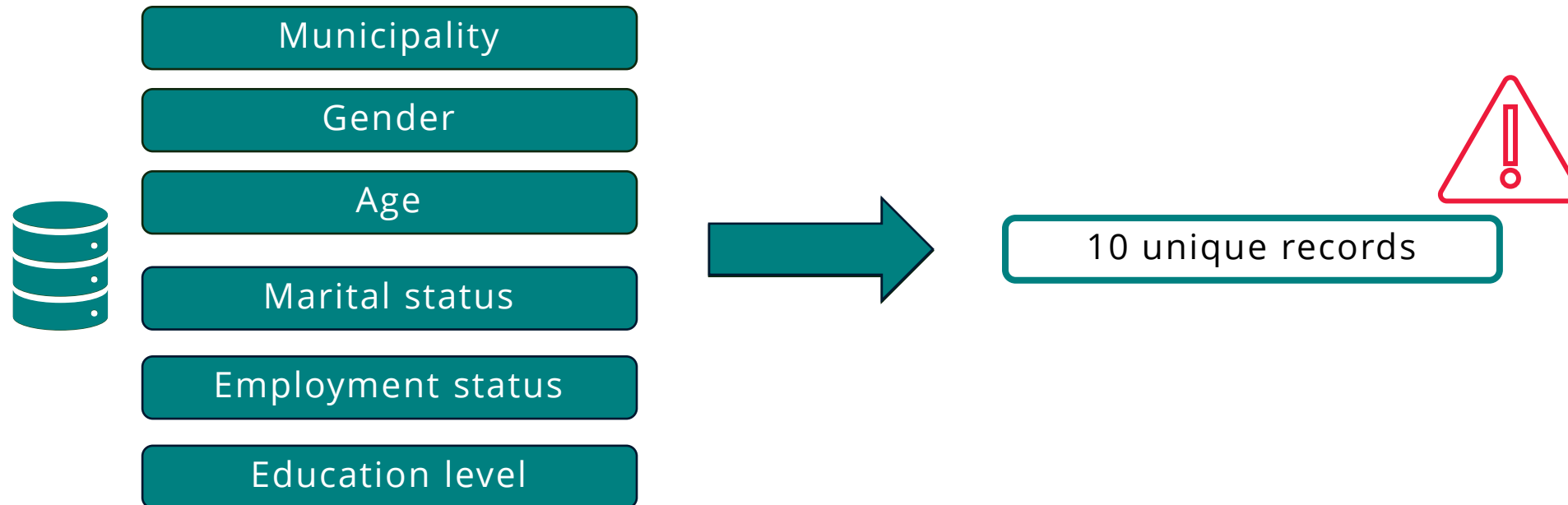
Unique identifiers enable the link between both datasets and the analysis of an integrated dataset





THE REIDENTIFICATION RISK ISSUE

A SIMPLE EXAMPLE



*But data protection may
become an issue...*

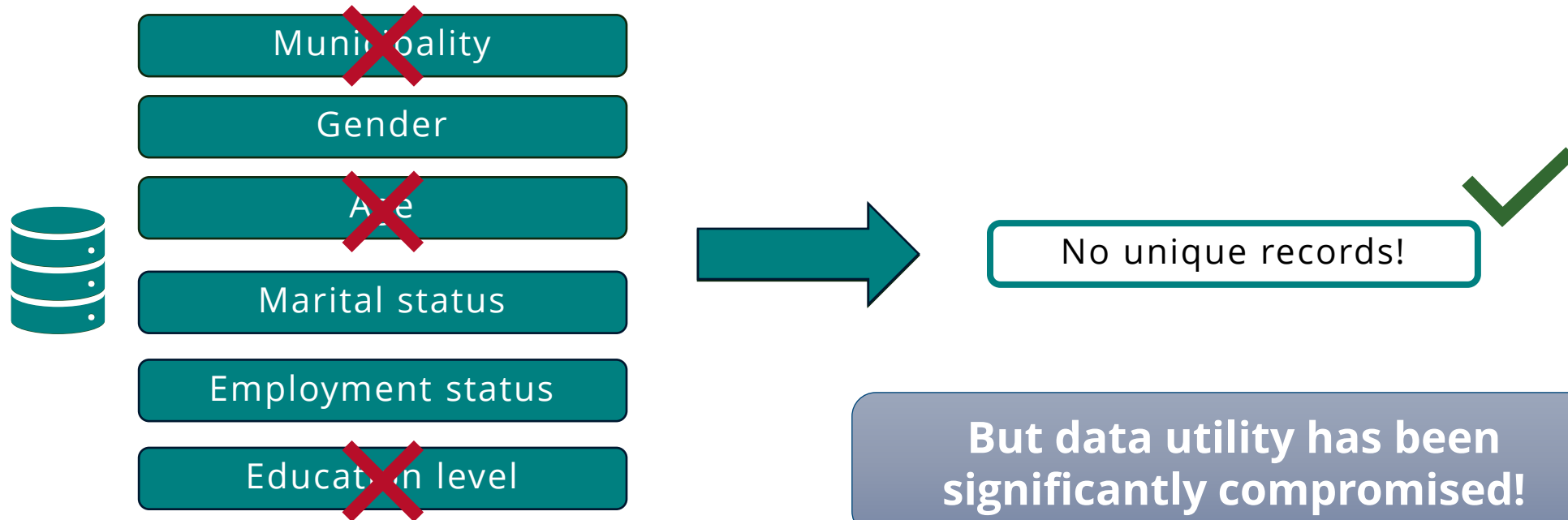




THE REIDENTIFICATION RISK ISSUE

A SIMPLE EXAMPLE

The easy way out...



But data utility has been significantly compromised!

Insights regarding age, location or education levels are lost

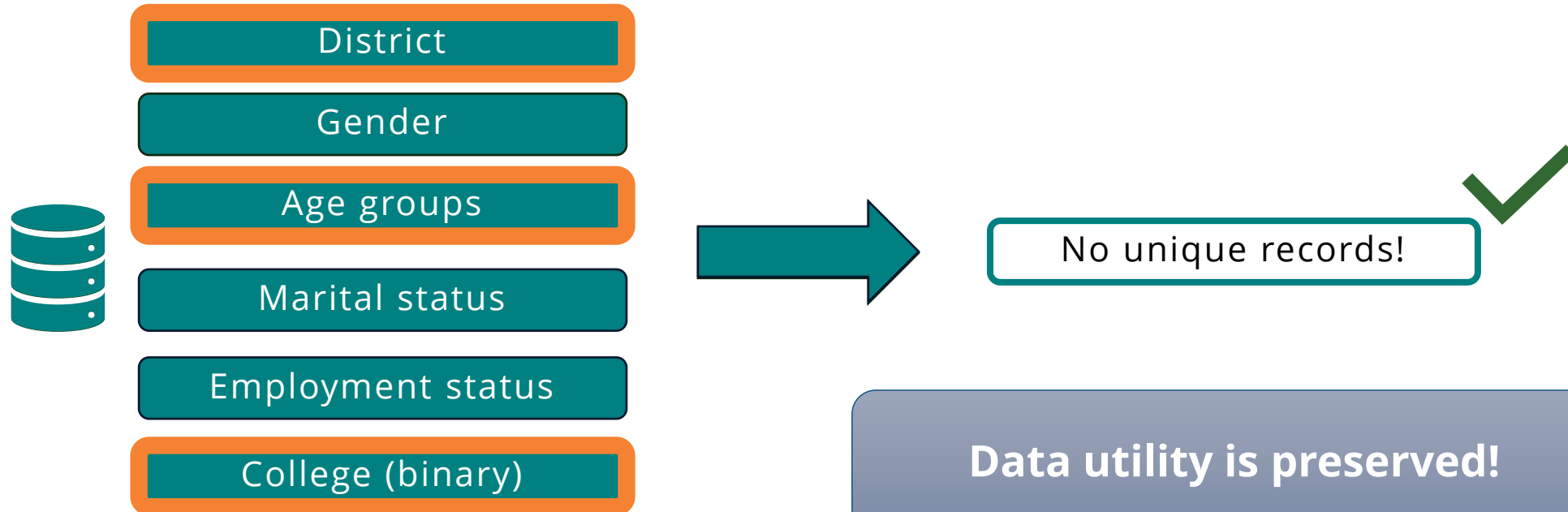




THE REIDENTIFICATION RISK ISSUE

A SIMPLE EXAMPLE

A more sensible approach...



*But this needs to be checked
whenever new data is available...*

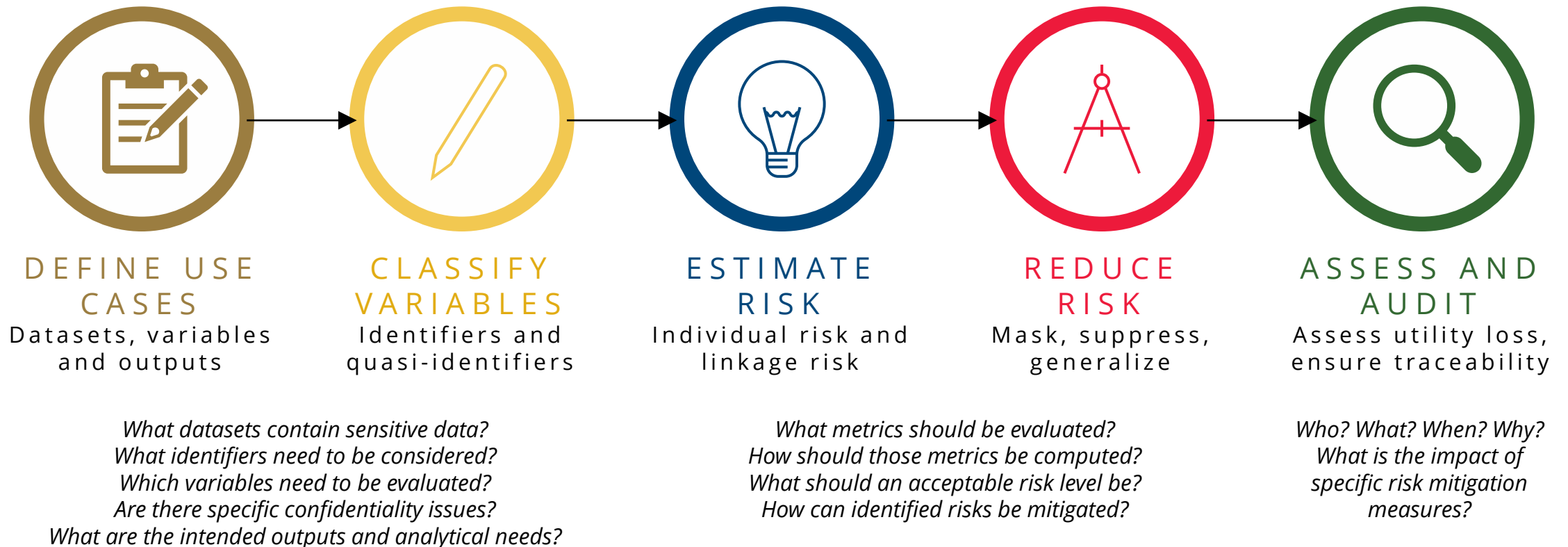




THE REIDENTIFICATION RISK ISSUE

A RISK-UTILITY ASSESSMENT FRAMEWORK

What steps need to be accounted to manage risk together with data utility?



The background of the slide is a photograph of a classical building facade, featuring a pedimented roof, columns, and ornate window decorations. A large, semi-transparent orange rectangle is overlaid on the image, covering most of the frame. The text 'TOWARDS OPERATIONAL IMPLEMENTATION' is written in white, uppercase letters in the top left corner of the orange area.

TOWARDS OPERATIONAL IMPLEMENTATION

04



TOWARDS OPERATIONAL IMPLEMENTATION

WAY FORWARD

The case-by-case framework...



Accounts for linkage risk



Minimizes information loss



Adapts to evolving risk



Harder to govern/operate... But this can be mitigated!



TOWARDS OPERATIONAL IMPLEMENTATION

WAY FORWARD

We are taking steps towards the implementation of a **harmonized anonymization/pseudonymization model**

This is particularly relevant as we are enhancing our **data integration strategy**

IReF

BdP DataHub

**The challenge is not choosing between privacy and utility,
but managing both simultaneously**



QUESTIONS

FFONSECA@BPORTUGAL.PT
MFLLOURENCO@BPORTUGAL.PT