



INSTITUTO NACIONAL DE ESTATÍSTICA  
STATISTICS PORTUGAL

# »» Organizational responses to multiple data collection - Administrative Data unit

Paulo Saraiva | Almiro Moreira | Sofia Rodrigues | João Poças ««



2022 UNECE Expert Meeting on Statistical Data Collection, Rome, 26-28 October

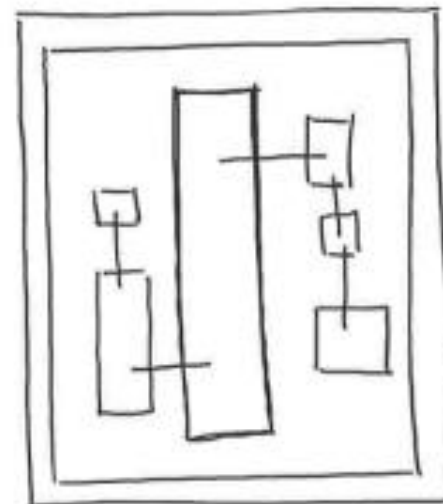
# Agenda

1. The National Data Infrastructure
2. The Administrative Data Unit
3. Examples

# 1. IND - National Data Infrastructure (1/4)

## Main Purposes:

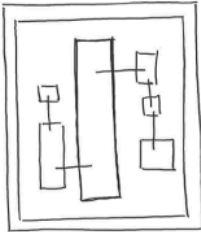
- Use of multiple data sources
- Increase the analytical capacity and flexibility in data management
- Exploitation of new data sources for statistical purposes
- Innovation



# 1. IND - National Data Infrastructure (2/4)

From 2018 onwards, the development of IND is ongoing:

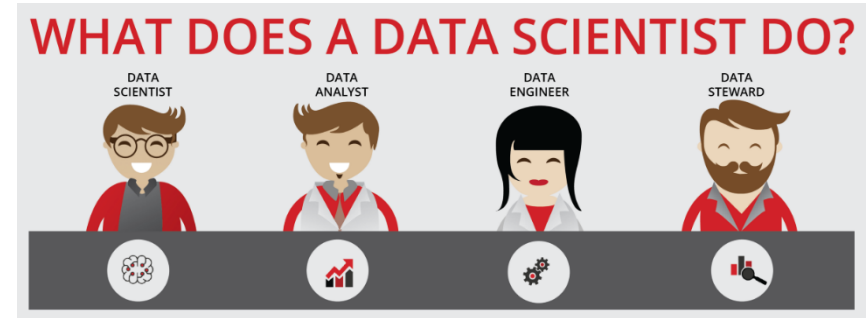
Single point access to several types of data and make it available to multiple purposes and users

Appropriation and use of administrative data and other sources	Partnerships	New roles
Data integration		Information security and data protection
Innovation	Integrate and share platforms, algorithms, knowledge, ...	Increase the economic and social value of the public good: statistical information

# 1. IND - National Data Infrastructure (3/4)

In 2019, adjustments in the internal organization for **strengthen** the capacity for data management and analysis in:

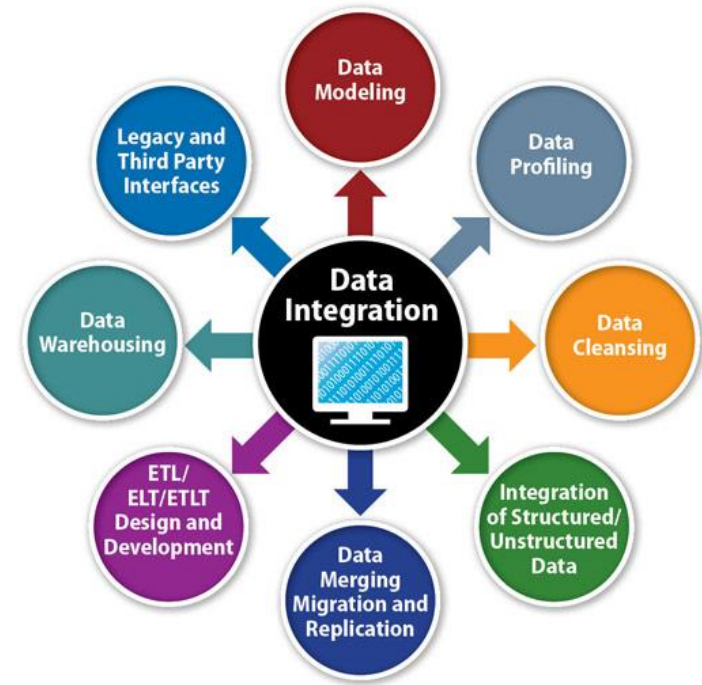
Methodology and Information  
System Department  
and  
Management and Data Collection  
Department



New skills, tools and techniques

## 2. The Administrative Data Unit

In 2020, creation of a dedicated Administrative Data Unit, under the Management and Data Collection Department



## 2. The Administrative Data Unit

### Main functions:

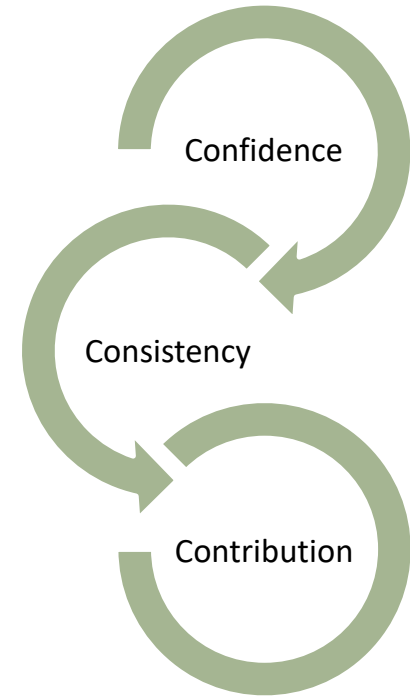
- Evaluation and testing the use of **new data sources**, with a view to improving the quality and consistency of statistical production;
- **Replacement** of the data collected by surveys or censuses;
- Definition of **new validation models**, consistency and coherence analysis;
- **Integration** of data from various sources
- **Simplification** the use of administrative data in the statistical production process

## 2. The Administrative Data Unit

This unit works in very close collaboration with Information Systems and Methodology Department, and all the other Subject Matter units.

It applies similar processes to different dataset ensuring greater data consistency.

It plays a very important role in the implementation of IND, contributing more to a data-driven society.





## 2. The Administrative Data Unit

Specific recruitment has been carried out for this unit, where scientific skills in terms of data appropriation, data analysis, programming skills and machine learning techniques are highlighted.

Up to now, this unit has **5 dedicated FTE**, and is in the process of recruiting two research fellows.

In addition:

- Hiring service providers specialized in the processing and analysis of large datasets;
- Specialized consulting contracts with academia;
- Specific training in programming languages for data analysis

### 3. Examples – e-invoice as a proof of concept

Under a protocol established between the tax authority and Statistics Portugal, monthly data on taxable values, derived from the e-Invoice system, are received since May 2020.

A huge amount of data is received every month: **around 80 million records.**

A set of **ETL procedures** are performed, introducing a set of significant improvements. Additionally, the database is also treated and enriched with:

- Inclusion of new attributes from other sources;
- Temporal referencing of characterization variables;
- Identification and imputation of anomalous values;
- Creation of new aggregated tables;
- Implementation of alternative plans.

### 3. Examples – e-invoice as a proof of concept

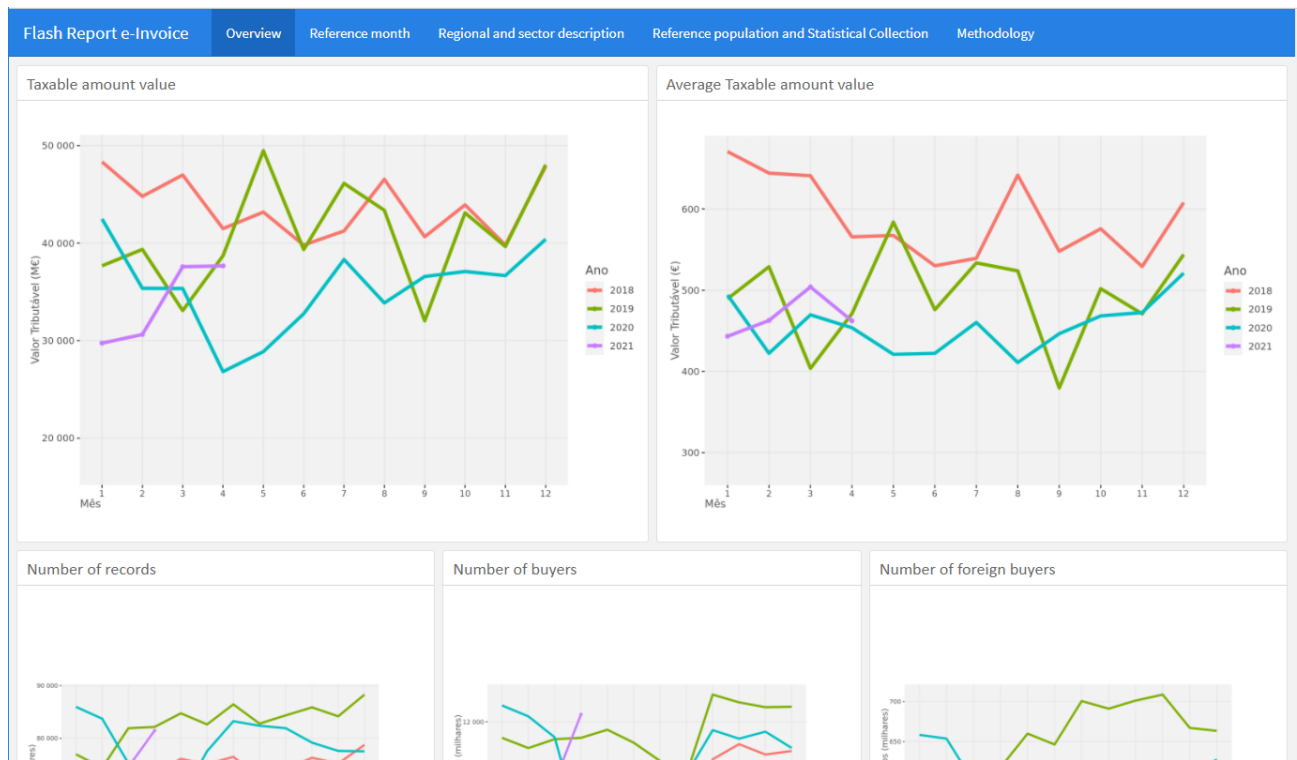
The e-Invoice data treatment played an important role as a proof of concept of the success of the Statistics Portugal administrative data strategy.

A set of procedures were defined that are being followed for all other sources, similar to what is done with traditional sources:

- methodological documentation;
- **flash reports (next slide);**
- definition of service level agreements.

# 3. Examples – e-invoice as a proof of concept

- Example of a Flash-report about e-invoice data produced every month
- Data and Flash-report delivered within **30** hours after data transmission



## 4. Conclusions

For all types of data, independently of collection mode (survey or admin)

**Key elements: standardization, integration, consistency**

Strong and centralized  
infrastructure

Dedicated Unit

**One consolidated database for all different users**



INSTITUTO NACIONAL DE ESTATÍSTICA  
STATISTICS PORTUGAL

## » Organizational responses to multiple data collection - Administrative Data unit

paulo.saraiva@ine.pt «

2022 UNECE Data Collection, Rome, 26-28 October

