Using Administrative Data for Statistical Purposes: Statistics Portugal Experience

Paulo Saraiva & Almiro Moreira
Statistics Portugal
Data Collection Department
Administrative Data Relevance

To preserve and improve **quality** …

…as well as reducing statistical burden and **costs**.
Administrative data for statistical purposes

Drivers to Relevance

• Higher awareness over excessive costs of context;
• Political actors → simplification of public services.
• Favorable environment → Administrative Data for statistical purposes.
Administrative data for statistical purposes

Critical Factors

- Metadata
- Stability
- Access
- Capacity (Statistical Authority)

Quality
Administrative data for statistical purposes

Frameworks: Europe and Portugal

**Portuguese framework**
- Law of the Portuguese Statistical System
  - 2008
- Organic Structure of Statistics Portugal
  - 2012

**European framework**
- Regulation on European Statistics
  - 2015
- European Statistics Code of Practice
  - P8 - Appropriate statistical procedures
    - 2011
  - P10 – Cost Effectiveness
    - 2011
- CoP Peer Review 2015
  - Recommendations 3 and 4
Mandatory Data Collection

“Statistical authorities may require the provision, free of charge, to all services or bodies, citizens or companies, of any information necessary for the production of official statistics”.

Prevalence over other special regulations

“These data provisions shall prevail over any limitations or obligations of secrecy contained in special domains, considering the use of administrative data for official statistical purposes as one of the determining purposes of their collection”.
Strong legislation is not enough

but... too many surveys.

Why?
Main Barriers to Access ADMIN

• **Cooperation** of sources and key governmental institutions;

• **Legal interpretation**: sources with specific rules not consistent with statistical legislation (e.g. Tax Authority);

• **Data security** is often an excuse to prevent the access;

• Lack of regular procedures to **monitor new legislative acts** and to **ensure the rights** to access;

• **Relevance** of Official Statistics: perception of its **usefulness**.
Administrative data are already used in several areas of official Portuguese statistical production:

- **Simplified Business Information**, essential for business statistics and National Accounts;
- **250 sources** benefiting **140 statistical activities**;
- Data Warehouse with **36 regular flows** and 1.5 billion historical observations.
<table>
<thead>
<tr>
<th>Flow</th>
<th>Periodicity</th>
<th>Calendar</th>
<th>Coverage</th>
<th>Update</th>
</tr>
</thead>
<tbody>
<tr>
<td>IRC – Business Income Tax - Annual Statement</td>
<td>Yearly</td>
<td>OCT y+1</td>
<td>2001 - 2015</td>
<td>17.09.01</td>
</tr>
<tr>
<td>IMI – Municipal Property Tax</td>
<td>Yearly</td>
<td>FEB y+1</td>
<td>2003 - 2015</td>
<td>17.02.01</td>
</tr>
<tr>
<td>Staff Statement - Employment and working conditions</td>
<td>Yearly</td>
<td>FEB y+2</td>
<td>2003 - 2014</td>
<td>17.03.01</td>
</tr>
<tr>
<td>ADENE - Certification of energy efficiency of buildings</td>
<td>Monthly</td>
<td>m+1</td>
<td>2013.10 - 2017.07</td>
<td>17.10.01</td>
</tr>
<tr>
<td>VAT - Value Added Tax</td>
<td>Monthly</td>
<td>2m after</td>
<td>2010.01 - 2017.06</td>
<td>17.10.01</td>
</tr>
<tr>
<td>DMR - Monthly Statement of Remuneration</td>
<td>Monthly</td>
<td>1m after</td>
<td>2014.01 - 2017.07</td>
<td>17.10.01</td>
</tr>
</tbody>
</table>
DW – Data Flow Technical Sheet

Administrative Data Flow – Technical Sheet

**Energy certification of buildings**

**Source:** ADENE - Energy Agency

**Description:** Energy certification and indoor air quality of buildings

**Periodicity:** Monthly

**Expected date of data availability:** Until the 15th of the month m + 1

**Data Warehouse Universe:** FNTADENE_TF_ADENE_20 (2014 onwards) and FNTADENE_TF_ADENE_10 (until 2013)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>COD_CERTIFICADO</td>
<td>Certificate ID</td>
</tr>
<tr>
<td>TIPO_EDIF</td>
<td>Building type</td>
</tr>
<tr>
<td>MORADA</td>
<td>Address</td>
</tr>
<tr>
<td>TIPO_IMOVEL</td>
<td>Property Type</td>
</tr>
<tr>
<td>ANO_CONSTR</td>
<td>Construction year</td>
</tr>
<tr>
<td>AREA_UTIL_PAV_TOTAL</td>
<td>Total floor area</td>
</tr>
<tr>
<td>TIPOLOGIA</td>
<td>Typology</td>
</tr>
<tr>
<td>EMISSOES_CO2</td>
<td>CO2 Emissions</td>
</tr>
<tr>
<td>CLASSE_ENERGETICA</td>
<td>Energy class</td>
</tr>
<tr>
<td>DATA_EMISSAO_CERTIFICADO</td>
<td>Certificate issue Date</td>
</tr>
</tbody>
</table>
IES is a form of electronic exchange of business information obligations of accounting, fiscal and statistical nature.

4 obligations

- Public entities
- Business legal obligations

Just once!
Statistics Portugal and Administrative data

Simplified Business Information

1. Tax Authority Portal
   - [Link to portal]
   - Complete Declaration, Submit Statement

2. SAF-T file

3. Ministry of Justice

4. Instituto dos registos e do notariado

Daily

48 h later

SAF-T PT (Standard Audit File for Tax purposes)

2017 Statistical Data Collection Workshop
Statistics Portugal and Administrative data

Simplified Business Information

BEFORE

50,000 companies (≈15%)
12 months
800 variables

AFTER

360,000 companies (≈100%)
6.5 months
5,000 variables

Business Survey
50,000 companies (sample) REPLACED
Statistics Portugal is preparing Census 2021 according to a new **mixed model**: administrative data with CAWI or CAPI interviews;

It is intended to increase quality and frequency, reducing burden and costs.
• 11 ADMIN sources evaluated (+100 million records);

• Of the 27 mandatory variables:

  *** 12 variables available in administrative sources and exhaustively characterize the universe;

  ** 6 variables available in administrative sources, but they partially characterize the universe;

  * 9 variables are not yet available in administrative sources.

Critical area: degree of relationship.
• Covered by the panel of indicators used in the alert mechanism under the Macroeconomic Imbalances Procedure;

• Calculated essentially with ADMIN data from tax authority, it is based on most transactions (quarterly).

Produced only with Administrative Data
Potential Sources Mapping Exercise 2015

- **Administrative Source**
- **Potential Source**
  - Can be used in the medium term
  - Can be used in the long term
- **Production Source**
  - Survey Simplification or Replacement
  - Validation
  - Registers
- **Legislative Changes**
Potential Sources Mapping Exercise 2015

- Systematization of the potential of Administrative data;
- Identified the potential of data available from public entities.

44 Potential Entities → 127 Relevant Sources → 106 Possible Sources → 8 Potential Sources
SIMPLEX (launched in 2006) is a government-wide program used to improve administrative and legislative simplification measures and to modernize public services.

Statistics Portugal participates in three measures:

- 2016 - Unique Information at Statistics Portugal
- 2017 - Single Information in Statistics Portugal+
- 2017 - Statistics Portugal: surveys Online
• **Measure**: to reduce the number and dimension of surveys, taking advantage of administrative data. Five Surveys: (1) **Turnover and Employment** (M), (2) **Labor Cost Index** (Q), (3) **Road Transport of Goods** (M), (4) **Intrastat** (M) and (5) **Housing Incomes** (Y).

• **Implementation**: 1st Quarter 2017

• **Main beneficiaries**: Privacy Policy | Companies

• **Sources**: Tax Authority and Social Security Institute

• **Status**: *Partially implemented* (limited replacement of the collection of the *turnover* variable for some companies).
• Measure: Continuing reducing the data requested by surveys: (1) 
  
  **Turnover and Employment: Industrial, Services and Retail Trade**, (2) 
  
  **Permanence of Hotel Guests.**

• Implementation: **1st Quarter 2019**

• Main beneficiaries: **Companies**

• Sources: **Tax Authority and Social Security Institute**

• Status: **In development** (e-invoice data to fully replace the collection of the **turnover** variable as well as **social security data** to replace **employment** and **remuneration** variables).
• **Measure**: Make available web interview (CAWI) as an alternative way to respond to surveys, in addition to face-to-face and telephone modes:
  - Survey on the **Use of Information and Communication Technologies by Families**.

• **Implementation**: 2nd Quarter 2018

• **Main beneficiaries**: Citizens

• **Status**: In development
Statistics Portugal and Administrative Data

Lessons Learned

- **Legislation** is important, but it's not enough;
- Assess **constraints** and identify **opportunities**;
- Make **practical results** known, even the small ones;
- Systematize the search for new **potential sources**;
- Benefit from **society's awareness of administrative costs**;
- Join government initiatives concerning service **modernization**;
- **Promote successful initiatives**. They are essential to increase the relevance of official statistics.
2017 Statistical Data Collection Workshop

Using Administrative Data for Statistical Purposes:
Statistics Portugal Experience

Paulo Saraiva
paulo.saraiva@ine.pt

Statistics Portugal
Data Collection Department