The National Data Infrastructure
New statistics and new partnerships

Francisco Lima

Conselho Superior de Estatística, 36ª Reunião Plenária

Lisboa, 15 December 2023

INSTITUTO NACIONAL DE ESTATÍSTICA
STATISTICS PORTUGAL

Outline

I. Statistics Portugal

II. The National Data Infrastructure

III. Partnerships

IV. Evidence

V. The way forward

I. Statistics Portugal

- Statistics Portugal is the central statistical authority,
- independent,
- with mandate for data collection and access to data,
- data for statistical purposes,
- producing official statistics with impartiality and objectivity

Statistics Portugal

Mission and Vision

The Mission of Statistics Portugal is to produce, in an **independent** manner, high-quality official statistical information, relevant for the society, while promoting the coordination, the analysis, the innovation and the dissemination of the national statistical activity and ensuring integrated data storage.

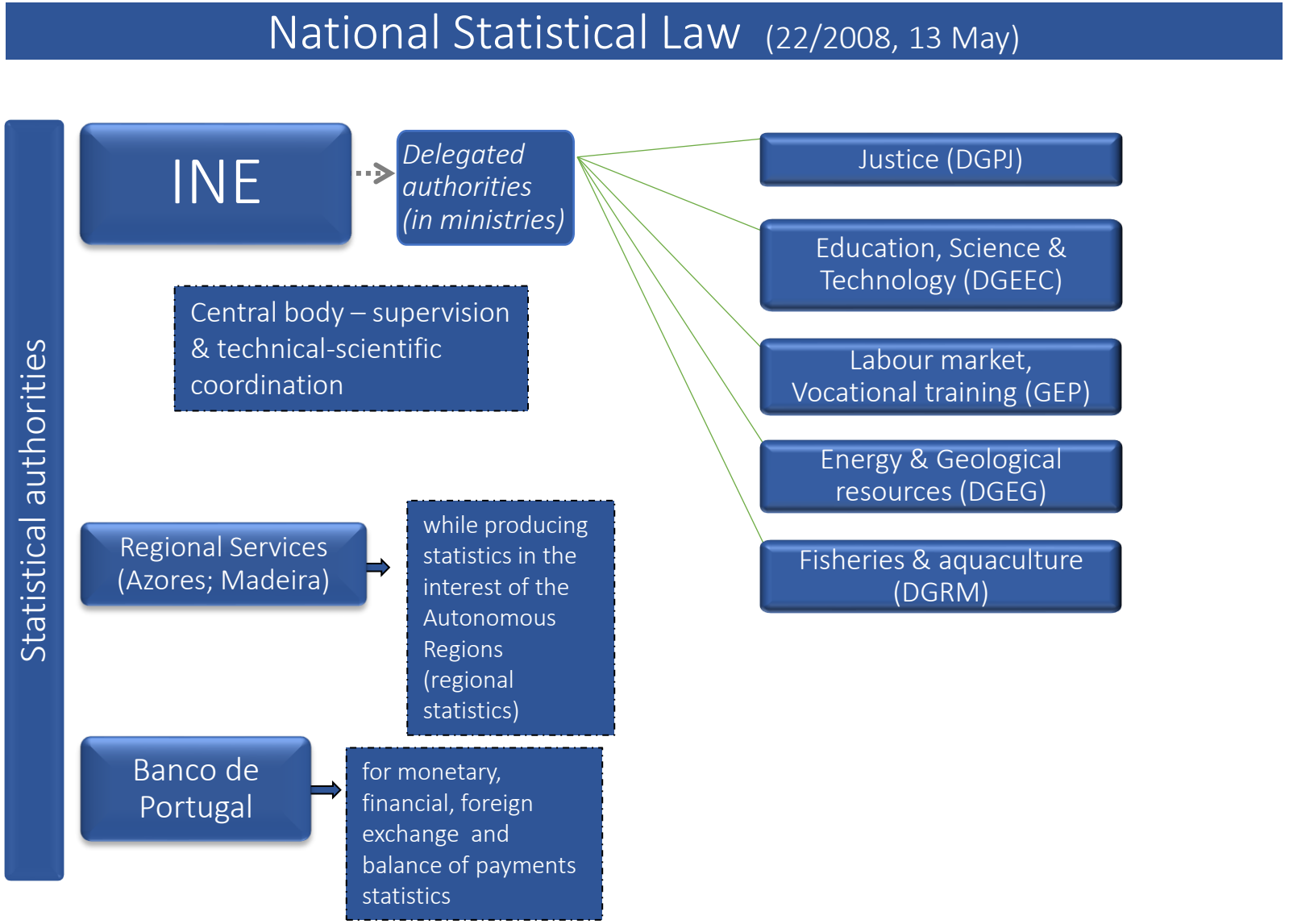
Vision. Statistics Portugal is recognized as an independent and trustworthy statistical authority, which develops methodological advanced statistical processes, resorts to technological innovation, and data science, and integrates multiple sources for statistical purposes. With all due respect to citizens and entities confidentiality, it delivers valuable statistics to society for better knowledge, research and decision-making.

Statistics Portugal

- Staff: ~600
- Delegations: Faro, Évora, Coimbra, Porto
  - Azores and Madeira with regional statistical services
- More than 770 information dissemination moments p/ year
- More than 370 press releases and publications p/ year
- The calendar rules
  - w/ a fixed daily hour

Statistics Portugal

A centralized system



International Statistical System

INE participates in different layers of decision and cooperation

European Statistical System (ESS)

- ESS Committee
- Eurostat + Member States National Statistical Institutes + EFTA + Switzerland

Community of Portuguese Speaking-countries (CPLP)

OECD

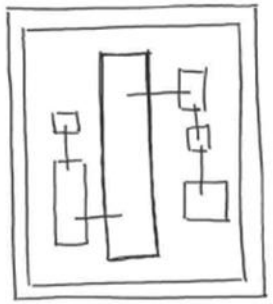
- Committee on Statistics and Statistical Policy

United Nations Economic Commission for Europe (UNECE)

United Nations

- Statistical Commission

II. The National Data Infrastructure

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

The National Data Infrastructure (NDI)

From 2018 onwards, the development of NDI by Statistics Portugal is ongoing

- Single point access to several types of data and make it available to multiple purposes and users
- Use of multiple data sources
- Increase the analytical capacity and flexibility in data management
- Exploitation of new data sources for statistical and scientific purposes
- Increase timeliness, relevance and granularity

The National Data Infrastructure (NDI)

- Data integration and interoperability
- Metadata catalogue – metainformation, **Data Catalogue**
- Confidentiality + Information security and data protection
  - **Certification** ISO 27001 (expanding), 27701 (planned), in addition to ISO 9001 (in process)
- Internal organization – assign new roles and create new units
- Specialization and increased complementarity between the different units
- Access other data – privately-held data (mobile, smart meters, IoT,…)

III. Partnerships

- Data providers
- Data users
- Academia
- Multidimensional partnerships

Partnerships with data providers

- Written agreements signed (or being prepared) with several public institutions that provide administrative data
- The agreements establish what's being provided and its frequency
- All agreements supported by a joint Technical Commission between Statistics Portugal and the partner institution
- Examples of institutions: Institute of Registries and Notary, Tax Authority, **Social Security**, **DG Health**, (ex)-SEF (Migration), …
- Evolve to a model, as in the agreement with the **Tax Authority**, where Statistics Portugal becomes responsible for providing data access to researchers and to other institutions

Partnerships with data users

- Fulfill the data needs of users: Media, public and private institutions (e.g., CCB, Fundação José Neves, Fundação, Francisco Manuel dos Santos, Infraestruturas de PT,…)
- Universities: U. Lisboa, ISCTE
- Users also use our APIs to extract directly information from the website

Partnerships with Academia

- Partnerships with Academia assume different forms: joint projects, grants to Master and PhD students, consultancy, internships for master students, training, …
- ISEG, IGOT, ISCTE, U Minho, UCP, IST, NOVA-SBE, NOVA-IMS, U Porto, …

Multidimensional Partnerships

- Statistical Regional Services – SREA and DREM
- Entities with delegated functions from Statistics Portugal – GEP, DGEEC, DGPJ, DGRM, DGEG
- Banco de Portugal (the central bank)
- DG Territory, Ministry of Health, Turismo de Portugal, Ministry of Agriculture, Portuguese Environment Agency (APA), Development and Cohesion Agency (AD&C), ANACOM…
- PlanAPP
- Municipalities, AML, AMP, CIMs, CCDRs

Access to microdata for scientific purposes

- Researchers from different institutions can access microdata for scientific purposes
  - The Integrated Business Account System
  - Quadros de Pessoal
  - Enterprises surveys – Innovation Survey, ICT use, International Trade, Management Practices,…
  - Family surveys – Labour Force Survey, ICT use, Survey on Living Conditions,…
  - Administrative data – IRS, IMT, IMI, IVA, e-fatura, …
- Can also access Public Use Files (PUF)
  - Census (5% sample), Agricultural Census,…

IV. Evidence

- Resident Population Database
- Integrated Data - Individuals and families
- Integrated Data – Enterprises
- Effort rates with permanent home loans
- SiT - Integrated Tourism Information System

Resident Population Database

- Main project of data integration
- The Census with Administrative Data
- Goal: To produce census statistics on an annual basis after the Censos 2021
- Comparison with the Resident Population Estimates points to <0.5% differences
- Challenge: move to a new model of population and housing statistics
- New opportunities for the development of social and demographic statistics

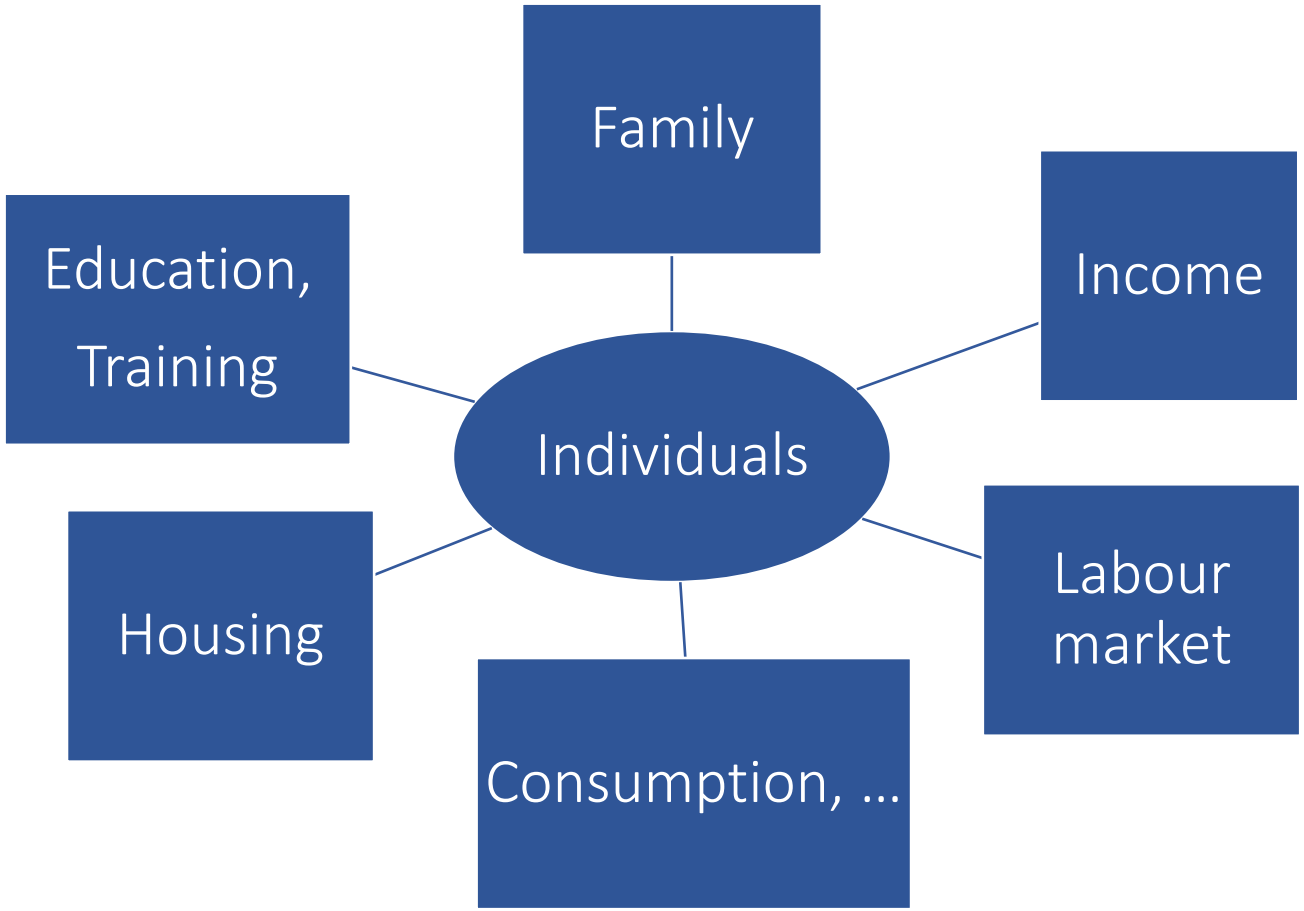
Resident Population Database

Major project of data integration

Instituto dos Registos e do Notariado **Civil Register**	Serviço de Estrangeiros e Fronteiras **Immigration Register**	Instituto de Informática da Segurança Social **Social Security Register**
Gabinete de Estratégia e Planeamento, MTSS **Private Employment Register**	Direção Geral de Estatísticas da Educação e Ciência **Education Register**	Autoridade Tributária **Tax register**
Caixa Geral de Aposentações **State Pension and Work Fund Register**	Instituto de Emprego e Formação Profissional **Unemployment Register**	DG de Proteção Social Trabalhadores em Funções Públicas **Social Protection for Public servants**
Instituto do Emprego da Madeira **Unemployment Register II**	Direção Regional do Emprego dos Açores **Unemployment Register III**	Direção Regional de Educação da Madeira **Education Register II**

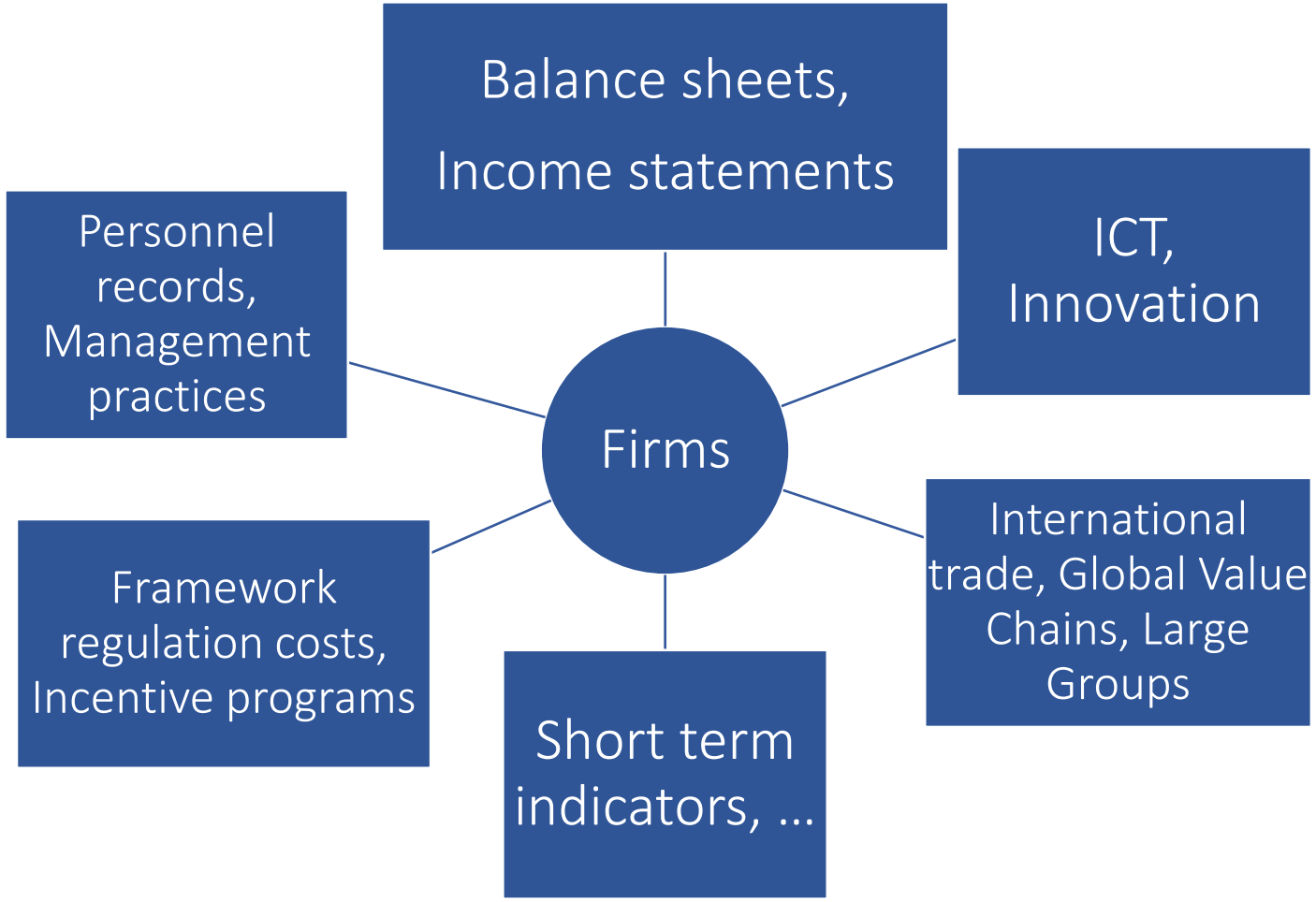
Integrated Data

Individuals and families



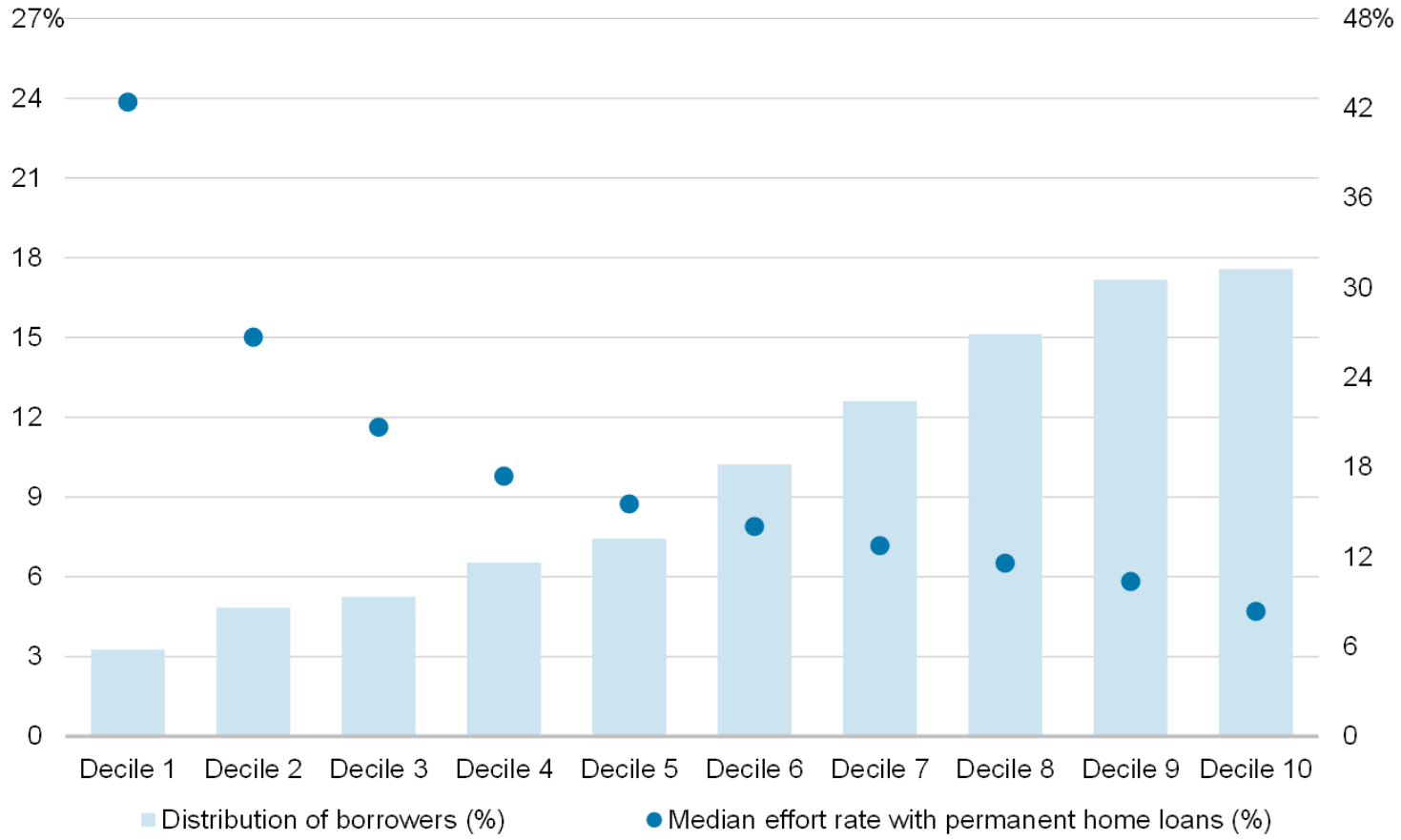
Integrated Data

Enterprises



Effort rates with permanent home loans

Distribution of borrowers (left axis) and median effort rate with permanent home loans (right axis) by decile of gross reported income less personal income tax paid , Portugal, December 2021



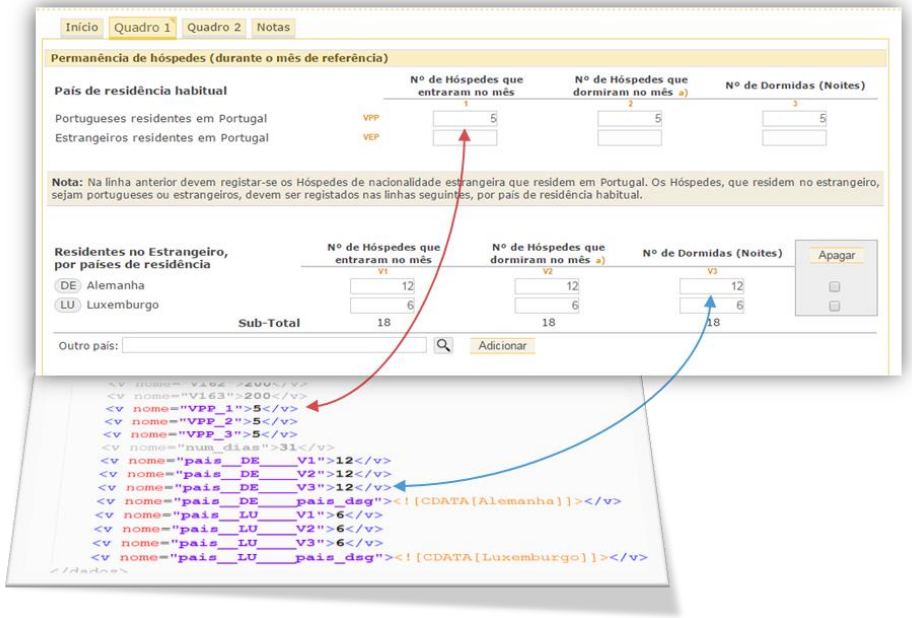
Effort rates with permanent home loans

- Releasing the effort rates with permanent home loans, referenced to 2021
- Integrate information from INE and the Central Bank (BdP)
- INE: advantage of the development of the IND - National Data Infrastructure - Income Statistics at local level (w/ IRS from the Tax Authority) and the Resident Population Database
- BdP: the Central Credit Register, which compiles information on credit instalments for the acquisition or construction permanent owner-occupied housing for borrowers with responsibilities to the national financial system

SiT Integrated Tourism Information System

- "**System to System**" approach and the adoption of a mixed model for data compilation, through administrative data appropriation and primary data collection

- **In-house development** and system supported by Statistics Portugal infrastructure – our *WebInq* platform (a single-entry point): primary data collection web page and automatic data transmission by XML



SiT Integrated Tourism Information System

- Shift from a multiple data collection model, where several entities collect similar information, to a **centralized data collection** and distribution system

- Ensure the **legal provisions** to serve the information needs of the project partners - the national tourism authority and the central bank

- Set of variables to be surveyed subject to **public consultation**, by the project partners, business associations and companies in the sector, as well as by software houses

- Except for the webservices to obtain administrative data, the **entire process will be guaranteed by Statistics Portugal**, from data collection to distribution to partners

SiT Integrated Tourism Information System

- A good example of what are the main goals of the National Data Infrastructure
- Statistics Portugal leverages its experience and in-house tools, and from the use of a widespread hierarchical language, making the sector transition and adaptation to the new transmission model much easier and minimizing the need for software development among most enterprises and software houses – **structural interoperability**
- Reduce burden and costs, eliminate redundancies, improve coverage, granularity and timeliness
- The ultimate output is a deeper knowledge of the sector to support policy evaluation and decision making

V. The way forward

- The will to move forward – the strategy
- Internal organization and organizational culture
- Human resources and human capital
- Data integration, interoperability, metainformation
- Political commitment and legal process
- Strong partnerships
- Communication with stakeholders – the benefits for society must be clear

Thank you

Francisco Lima

INSTITUTO NACIONAL DE ESTATÍSTICA
STATISTICS PORTUGAL