

## Data collection Dashboard: a step towards more consistent data

João Poças; Alexandre Cunha; Almiro Moreira; Paulo Saraiva; Sofia Rodrigues (Statistics Portugal)

[joao.pocas@ine.pt](mailto:joao.pocas@ine.pt); [alexandre.cunha@ine.pt](mailto:alexandre.cunha@ine.pt); [almiro.moreira@ine.pt](mailto:almiro.moreira@ine.pt); [paulo.saraiva@ine.pt](mailto:paulo.saraiva@ine.pt);  
[sofia.rodrigues@ine.pt](mailto:sofia.rodrigues@ine.pt)

### *Abstract*

Analysing and ensuring the consistency of data from different sources, either internal or external, is one of the most important aspects for producing quality and consistent statistics.

Within the scope of this analysis, a daily Dashboard was developed containing a set of statistical units with relevant economic activity, which allows an analysis not only between external data sources (administrative) but also integrating data from internal sources (surveys).

This internal production tool was developed in R Markdown and is produced automatically and available on Intranet. It allows a comparison of several data sources in an interactive way, starting from a more general view to a more detailed view.

It contains not only a set of main indicators but also internal observations from contacts with respondents or from previous analyses, thus supporting faster analysis and decision-making, both within the scope of data collection (sharing information between surveys) and within the statistical production process.

# Data collection Dashboard: a step towards more consistent data

**Authors:** João Poças, [joao.pocas@ine.pt](mailto:joao.pocas@ine.pt); Alexandre Cunha, [alexandre.cunha@ine.pt](mailto:alexandre.cunha@ine.pt); Almiro Moreira, [almiro.moreira@ine.pt](mailto:almiro.moreira@ine.pt); Paulo Saraiva, [paulo.saraiva@ine.pt](mailto:paulo.saraiva@ine.pt); Sofia Rodrigues, [sofia.rodrigues@ine.pt](mailto:sofia.rodrigues@ine.pt); Statistics Portugal

## 1. Collecting data from different sources

The focus of statistical offices on the collection and use of external data (namely administrative data) for statistical production has been a growing trend in recent years.

In order to strengthen and proactively participate in this trend, Statistics Portugal has recently taken some initiatives and made some adjustments in its internal organization:

- As of 2018, the development of the National Data Infrastructure (NDI) is underway, with various objectives such as: integrating and using multiple data sources, increasing analytical capacity and flexibility in data management, exploiting new data sources for statistical purposes;
- In 2019, the roles of Department of Methodology and Information Systems (DMSI) and Department of Data Collection and Management (DRGD) were strengthened in their responsibilities of data management and analysis;
- In 2020, a dedicated unit devoted to treatment and analysis of administrative data sources was created, within DRGD: The Administrative Data Unit.

These changes, both in terms of organizational adaptations and in terms of hiring and training human resources with analytical, statistical and technological skills, were essential to promote a closer relationship with external data providers while aiming to facilitate, internally, the integration of new data sources in statistical production.

## 2. Issues when integrating different sources

Statistical production has always sought different forms of data collection and different data sources to improve the quality of the statistics produced. Traditional data collection (through surveys) has increasingly incorporated data from external sources (e.g. administrative data) into its processes and quality analysis.

However, the use of data from different sources always presents new challenges and difficulties in identifying and validating the most accurate data and thus eliminating possible inconsistencies.

Faster analysis and evaluation of inconsistencies makes it possible to improve data quality, speeding up statistical production and, at the same time, reducing possible contacts with respondents (and therefore reducing the statistical burden).

### 3. The creation of the Data Collection Dashboard

Considering that the ease and speed of analysis are essential factors for a higher quality and faster statistical production, the idea was to develop a tool combining indicators available in various data sources, whether internal or external, with the highest possible timeliness.

Thus, and without the need to resort to the IT development unit, the idea arose to build a Dashboard that was not only intuitive but also sufficiently appealing and updated frequently (daily basis), in order to assist, on one hand, the work of the data collection units and, on the other, the statistical production units. The tool chosen to implement this Dashboard was the R language, more specifically using Flexdashboard technology, since there was already some know-how in the newly created administrative data processing team in the data collection department.

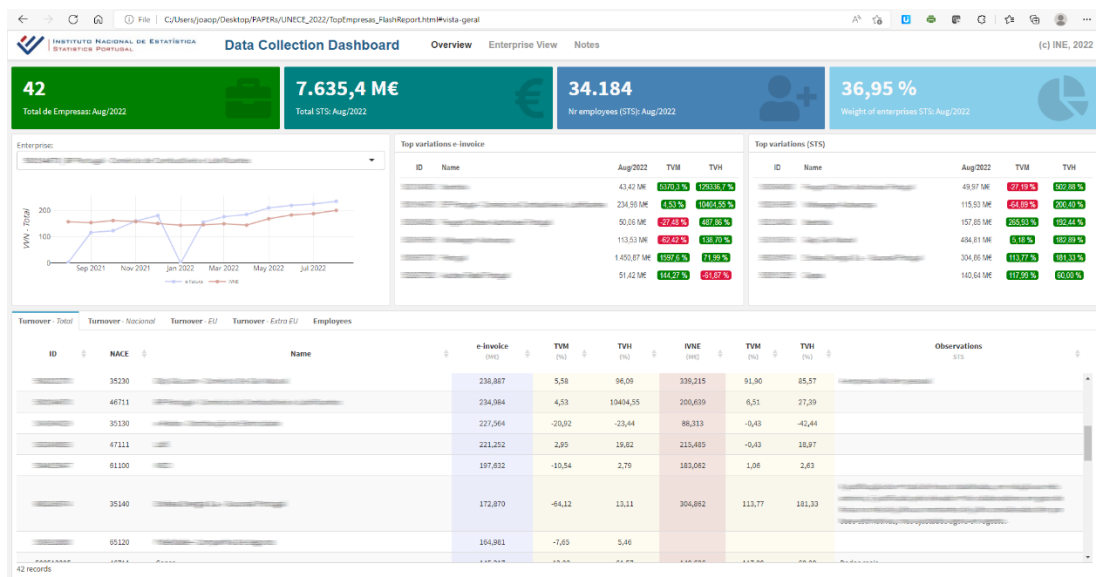


Figure 1 – The “overview page” of the Data collection Dashboard

### 4. Dashboard contributions to overcome inconsistencies

This Dashboard is therefore an up-to-date overview of a relevant set of economic statistical units, on the current reporting of some key indicators, from different data sources (surveys and administrative data) as well as aggregating other elements that may explain certain behaviors in the indicators and thus avoid more detailed and time-consuming analyses with eventual contacts with respondents.

Specifically, with the implementation of this Dashboard, analysis can begin with a more aggregate view, indicating the aggregate values of some main indicators, which statistical units have the greatest variation (year-on-year and quarter-on-quarter) and, for the same reference period, comparing these values both in terms of data collected via survey and administrative sources. It is also possible to provide a more detailed perspective for each unit, which goes beyond the simple provision of the time series, also using the indication of any observations recorded by the data collection teams, when any unusual behavior is detected.

The information made available in this way and updated on a daily basis is of great importance for the teams that have to perform quality analysis of the data collected on a daily basis, allowing them to have both an overall and specific view of certain key indicators.

## **5. Conclusions**

The development of the National Data Infrastructure (NDI) is ongoing, since 2018, with a vision of integrated and consistent data that allow increasing analytical capacity and statistical production.

In recent times, Statistical Portugal has made a significant investment in promoting, among its staff, the learning new skills, tools and techniques, in order to overcome the difficulties inherent in processing a large amount of administrative data, to be made available in a very short space of time to internal users.

In this context, the implementation of this Dashboard plays an important role not only in helping to collect and treat data in a more consistent way, enhancing the improvement of its quality, but also in promoting collaboration among the various teams of the statistical process.

This solution allows us to think about creating other dashboards for other reference times/frequencies using other data sources. Additionally, the same principle can be applied in creating Dashboards for specific analyzes (eg SPE - Special Purpose Entities)