



Economic and Social Council

Distr.: General
7 June 2023

English only

Economic Commission for Europe

Conference of European Statisticians

Seventy-first plenary session

Geneva, 22–23 June 2023

Item 9 of the provisional agenda

Timeliness, granularity and frequency of official statistics

How to keep timeliness and frequency with monthly administrative data?

Prepared by Portugal

Summary

The increasing and intensive use of non-statistical data sources (administrative and others) brings important challenges to the official statistics production process. Statistics Portugal has developed in the last years several initiatives aiming to respond to these challenges, in which we highlight the creation of the National Data Infrastructure, a strategic objective reinforced by several structural changes: adjustments in the internal organization to strengthen the capacity for data management and analysis; the creation of the Administrative Data Unit, devoted exclusively to the collection and analysis of administrative data with an emphasis on the integration of different data sources; and the updating of the Statistical Production Process Handbook. Complemented with significant investments in learning new skills, tools and techniques to process and analyze (massive) sets of data, contributed to the capability to produce timely statistics. E-invoice data (60-90 million invoices per month from the Tax Authority) present a proof of concept of this strategy. However, it also raises issues concerning consistency and sustainability, given the dependency on the data provider. The speed – less than 48 hours – to deliver a large set of already processed data, transformed and enriched with other data sources, proved to be an important contribution to the construction and achievement of the objectives of the National Data Infrastructure.

This document is presented to the Conference of European Statisticians session on “Timeliness, frequency, and granularity of official statistics” for discussion.



I. The framework and Statistics Portugal approach

1. How can national statistical institutes (NSI) give more to society? How to produce more, meeting the growing demand for relevant, granular, and timelier statistics? And facing lack of scale, human resources restrictions and the need to constantly innovate? When getting increasing access to data? Embrace other functions, sharing our expertise in metadata, statistics, algorithms, and platforms, while keeping the quality of the regular production? Too many questions raising important challenges to official statistics. Each NSI (and international organizations) devised its own response to these challenges. In this vein, Statistics Portugal has developed in the last years several initiatives, in which we highlight the creation of the [National Data Infrastructure](#).

2. The main objective of the National Data Infrastructure is to adopt a more intensive and integrated use of administrative data in the production of statistical information, taking advantage of the whole production chain of Portuguese official statistics. This chain ensures the protection and integrity of the data, from the development of platforms, applications and algorithms, data collection and validation to the analysis of the statistical information. In practical terms, we intended to create a single point of access to the various types of data and make it available to serve multiple purposes or projects, either to produce official statistics by Statistics Portugal or for research purposes.

3. Several structural changes that complement each other reinforced this strategic objective. Adjustments in the internal organization strengthen the capacity for data management and analysis. The creation of the Administrative Data Unit, devoted exclusively to the collection and analysis of administrative data with an emphasis on integrating different data sources, can also be considered highly relevant and innovative¹. The results already benefited several processes with the integration of different sources and the potential benefits are vast.

4. The Statistical Production Process Handbook, a reference in our production system, was also updated in 2020 to, among other reasons, introduce the National Data Infrastructure project in statistical processes through its mention in specific tasks and to put a greater emphasis on appropriation, validation, treatment, analysis, and integration activities of microdata derived from alternative sources (other than surveys). Complemented with a reinforcement of partnerships with data providers, data users and the academia, in addition to significant investments in learning new skills, tools, and techniques to process and analyze (massive) sets of data, all these changes contributed to the capability to produce timely statistics.

II. The e-Invoice as a proof of concept

5. Access to administrative data is being established through collaboration protocols specifying the registration design, transmission dates, and information on security measures, complying with the provisions of the National Statistical System law and the General Data Protection Regulations.

6. Statistics Portugal established a written agreement with the Tax and Customs Authority covering a significant and varied number of data transmissions. One of the data sets consists of the information from electronic invoices (e-Invoice), a reporting invoices system implemented by the Tax Authority for tax purposes (VAT) and as part of the administrative simplification and anti-fraud measures. Electronic transmission of the invoices issued by individuals or legal entities with their head office or permanent establishment in Portuguese territory to the Tax Authority is mandatory. Statistics Portugal

¹ Especially in a country lacking a unique individual identification number across the whole public administrative system and still relying heavily on surveys for producing official statistics.

receives, monthly, 60 to 90 million observations with information on both the seller and the buyer².

7. The centralization of the treatment processes and the integration with other data sources improve the consistency and quality of data. A data driven approach that allows us to leverage economies of scale, instead of an approach based on subject matters, prone to duplication of activities, inconsistencies across the organization and missing opportunities for statistical development. Working closely with Methodology and Information Systems Department and with subject matter units, the Administrative Data Unit prepares the information for internal users.

8. The integration with other data sources creates opportunities to develop and produce new statistics, more granular and timely, for example, on: business turnover (economic activity, namely at the regional level³ or to replace the variable in the monthly business turnover surveys); individual consumption (e.g., related to Covid-19⁴ or with the birth of the first child⁵); signs of life for the development of the statistical population register (and the administrative Census⁶); retail influence areas (e.g., pharmacies) or households' mobility patterns. The new statistical results obtained with e-Invoice (and with other data sources) are regularly published in [StatsLab](#) – an area in Statistics Portugal website dedicated to disseminate new statistics (not yet official) resulting from administrative and unconventional data sources and new methodologies, taking advantage of the potential of the information associated with Statistics Portugal's National Data Infrastructure.

9. The Administrative Data Unit also produces the metadata and monthly reports that include the main results and a list of the main characteristics of the e-Invoice data set and from the integration with other data sets. The speed – less than 48 hours – to deliver a large set of already processed data, transformed, and enriched with other data sources, proved to be an essential contribution to the construction and achievement of the objectives of the National Data Infrastructure.

III. Issues to consider – what can go wrong?

10. Although Statistics Portugal has signed a growing number of protocols with administrative data holders (e.g., Institute of Registries and Notary, Tax Authority, Social Security, Immigration and Borders Service), regular consultation and cooperation are still challenging. Some administrative processes are very complex and imply a deep knowledge of the many different events originating from the interaction of citizens and companies with the public administration. Although explanatory notes are usually associated with administrative forms, the available metadata might not be compatible with the requirements of statistical production.

11. Statistics Portugal shares the results and achievements on the use of administrative data with data providers. Still, in general, there is some difficulty in understanding the very demanding official statistical needs. The GDPR poses to all NSI some additional challenges concerning access to personal data that, in some cases, might limit the access to certain statistical domains, even though the principle of confidentiality under the current national law already gives the protection needed by personal data holders.

12. Concerning e-Invoice in particular, the transmission of data does not occur with the regularity of statistical production needs concerning the number of observations and, of course, its correspondent sales value, depending on the day of the month that the Tax Authority provides the data. Two or three days earlier in a month can significantly impact

² Regarding the protection of personal data, given its sensitive nature, Statistics Portugal applies technical and organizational measures to its processing that aim to guarantee security, confidentiality, integrity and availability. These measures are characterized using irreversible encryption algorithms that protect sensitive attributes, particularly the unique identifiers of records, and in the application of other methods that allow to avoid the direct or indirect identification of individual entities.

³ [COVID-19: How has the pandemic affected regional economies in 2020? \(ine.pt\)](#)

⁴ [Expectations, Infections, and Economic Activity \(nber.org\)](#)

⁵ [Birth of the first child: what changes in the pattern of family expenses? \(ine.pt\)](#)

⁶ [Census with administrative data \(ine.pt\)](#)

the information transmitted. As the data source becomes increasingly integrated into the regular production process (and even to replace survey questions), the dependence on the administrative data providers is a critical issue.

13. To address this issue and concerning the e-Invoice, Statistics Portugal is working on three different fronts:

- (a) Setting up an automated procedure to identify and fill in any potential missing data and outliers, for the most relevant issuers (sellers);
- (b) Negotiating with the Tax Authority to submit additional versions of the same reference month, allowing to develop a more robust and comprehensive estimation model;
- (c) Defining an alternative approach, a “Plan B” (e.g., nowcasting), in the case of (serious) problems in the data transmission.

14. Despite the dependency on the data provider, namely in ensuring consistency and sustainability of the data transmission, the use of administrative and other data, namely privately held data, is a path of no return, given all its benefits and advantages. It is up to the NSI to effectively convey the message and demonstrate the benefits regarding the importance of using multiple sources in the production of official statistics, along with the need for accurate and reliable data, while ensuring alternative plans.

15. Internally, the change in the organizational structure and the production process also raises some challenges. The different units becoming more interdependent are asked to change gradually the working relations and engage in a permanent dialogue as they strengthen their complementarity. Creating units with statisticians specialized in data management and engineering – collection, validation, integration, treatment – implies that the subject matter units should focus more their efforts on the development of new statistics and statistical analysis. In some cases, this implies a change in the organizational culture that may take some time to implement.

IV. Conclusion

16. The case presented shows how we have been dealing with the challenges faced by official statistics in the past years. It is not the full story but addresses some components of our response to the need for new, timelier, and detailed statistics.

17. Taking advantage of the robustness of Statistics Portugal framework, including its development of the National Data Infrastructure, the future perspectives are based on a higher level of data integration of multiple data sources, including surveys, censuses, administrative sources and privately held data. Working on integration, consistency, coherence and quality allows to create value and expand the frontier of production possibilities, guaranteeing the production of official statistics and offering new statistical products. Besides the vital role that official statistics already play in the making of public decisions by providing objective and reliable information about various aspects of society and the economy, Statistics Portugal envisages to have a stronger role in the data ecosystem, using new methods and tools concerning processing, validation, enrichment, anonymization, protection and data security, contributing to a more widespread use both of data and of statistics produced.

18. Timeliness, frequency and granularity impose a strong focus on securing access to data sources, investing in skills and promoting innovation across the entire organization. To look for opportunities created by new technological developments that can improve our work, particularly a more intensive use of artificial intelligence, data protection methods and new IT tools. Strong partnerships with data providers, users and the academia are likewise essential. Effective communication approaches on how relevant these external stakeholders are in the community of official statistics should also be enhanced. We must also be willing to change the organizational structure and improve the statistical production process, even if it causes some perturbation of the usual way of working and challenges the existing organizational culture.