Tourism Data: Integrated Information System (S2S), sharing data and Official Statistics
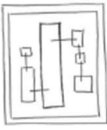
Rui Martins, Sofia Rodrigues, Maria Jordão and Carla Braga [1]

(1) Statistics Portugal

1. Introduction

In line with the strategy of strengthening statistical production through the appropriation of administrative data, Statistics Portugal (INE) has developed several initiatives in recent years, among which the creation of the National Data Infrastructure (IND) stands out. Its main purpose is to provide a single access point to the various data sets (administrative and other) and make them available for multiple purposes or projects, either to support the production of official statistics by Statistics Portugal or for research activities.

In this strategy, tourism is a very relevant statistical area to develop a project of this kind, for several reasons. It is an activity within a very dynamic sector, highly competitive, with a strong local, regional, national, and international impact, and with a growing importance in the country's economic activity.

The development of the Integrated Tourism Information System (SiT) is an example of what is being done in the scope of the IND, regarding both data sharing based on a "System to System" approach, and the adoption of a mixed model for data compilation, through administrative data appropriation and primary data collection. The whole development is carried out in-house and the system is supported by INE infrastructure, in particular, in the use of WebInq platform: primary data collection web page and automatic data transmission by XML.

This system main goal is to deliver statistical data through a single-entry point - WebInq -, and subsequently share the relevant information with the project partners, which are, in this case, the national tourism authority - Turismo de Portugal (TP) - and the national central bank - Banco de Portugal (BdP). This system, with one single interaction, is expected to generate advantages both for enterprises, easing the administrative burden on them, and the main stakeholders and partners involved in the project, which will have access to more useful and timely information. Moreover, this system will allow for the production and release of new indicators and dissemination products, contributing to a more in-depth knowledge on tourism activity in Portugal.

2. The data transmission system

The development of SiT will allow for a significant efficiency gain, shifting from a multiple data collection model, where several entities are collecting similar information and thus with some redundancy, although in different formats, to a centralized data collection and distribution system.

The SiT development is being based on the reporting of statistical information through the automatic data transmission, via XML (through upload or webservice), with electronic forms being designed according to the different characteristics of the establishments and their configuration.

This solution is often used in the scope of surveys with a low frequency, like weekly or monthly, or surveys whose respondents use third-party applications (e.g. software houses solutions) that are highly widespread and entities that have information systems containing data required by the NSI. Additionally, it's easier to implement when the entities belong to large business groups (with a high statistical burden) and handle large volumes of data in order to respond to surveys.

As advantages, we can expect a reduction in the number of contacts with respondents, less non-compliance situations with response deadlines, having the possibility of anticipating the response. Data is expected to be collected faster and with less costs.

Possible difficulties for respondents include the initial investment to adapt their information systems to TAD, the possible "purchase" of a TAD connection module available in third-party applications (e.g. software houses).

The solution adopted, already implemented in the current survey on guest stays in hotels and other establishments, is a widespread hierarchical language (XML (Extensible Markup Language)), with great freedom in creating the structure, easy to implement and maintain in evolutionary terms. It presents a simple implementation of validations to the structure (XSD files), regardless of the platform or software used, and is readable by applications and people.

The structure defined for the file also has the advantage of being used by multiple Entities, Surveys and Occurrences.

The information can be submitted in two ways: on INE's Webinq page, by sending the XML file, or via Webservice, in which case it can be called up directly by external systems, via secure channel communication (data encrypted via SSL certificate).

Although different, both solutions require the credentials of a WebInq service member (member code / password).

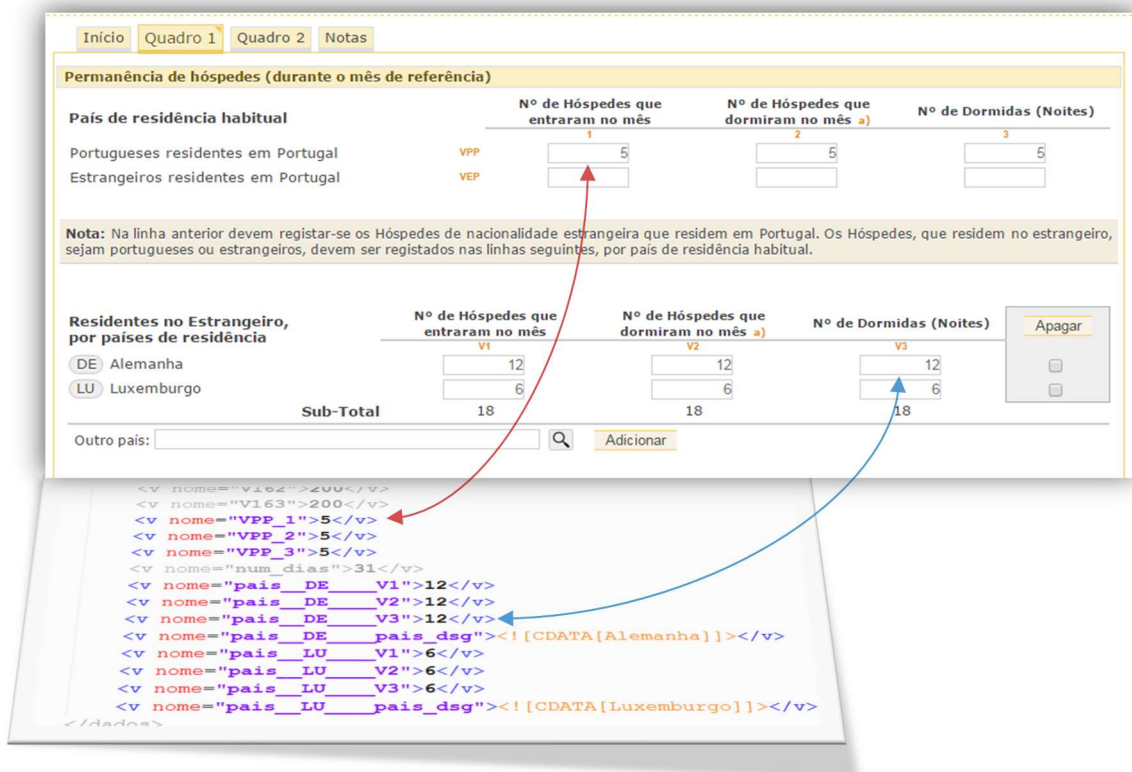Figure 1 – Scheme of a real XML to response IPHH

Figure 2 – Fill the webform with content of XML file

Except for the webservices to obtaining administrative data, the entire process will be guaranteed by Statistics Portugal, from data collection to sharing with partners, as well as the dissemination of statistical information.

3. Additional developments

Additionally, in the scope of the SiT, Statistics Portugal intends to start covering small Local Accommodation (with less than 10 capacity users), which is currently not mandatory and thus not covered in the collection of statistical information on tourism activity, despite its significant and growing presence in tourism activity in Portugal. This group of accommodations have very specific characteristics: the heterogeneity of the operators they include (from private individuals looking for a complementary economic activity to professional companies with the most modern management systems) and the volume of overnight stays they currently record. Electronic forms are being designed according to the different characteristics of the establishments and their configuration.

The upstream construction of SiT is based on administrative data collected from the National Tourism Registry (RNT) of the national tourism authority, which includes the registration of tourist accommodation establishments (ET) and local accommodation (AL). A link is established, via webservice, allowing Statistics Portugal to access daily data on establishments' licensing, such as new openings, closures, changes of categories, etc.. After the appropriate treatment, these registry data make it possible to move forward and the statistical information collection process is then carried out by Statistics Portugal.

Since 2021, the national tourism authority makes available to INE, via webservice, the new registrations data, making it possible to cover the opening of new tourism accomodation establishments in real time and to start collecting information on their activity in a much shorter period.

For this purpose, Business Register was adapted to the new variables from the RNT, previously carrying out a data cleaning and aggregation work of AL records, through the item ID matrix article, resulting in the new statistical units for future integration in the information reporting through SiT, in a less burdensome way for the respondents.

The set of variables to be surveyed was subject to public consultation, by the project partners, business associations and companies in the sector, as well as by software houses. Their contribution allowed for the identification of all information needs and the requested variables available for reporting.

4. Conclusions

The development of the Integrated Tourism Information System, despite raising some challenges that are still under study, is a good example of what are the main goals of the IND, allowing for efficiency gains, through easing the statistical burden on respondents, eliminating redundancies among the different stakeholders of the sector collecting information, while significantly improving coverage, granularity and data timeliness. The ultimate output from this is a deeper knowledge of the tourism accommodation activity in Portugal to support decision and policy making.

For developing this new system, Statistics Portugal takes advantage from its experience and inhouse tools developed, as well as from the use of a widespread hierarchical language, both making the sector transition and adaptation to the new transmission model much easier and minimizing the need for software development among most of the enterprises and software houses. Although the challenge is a bit different within the small local accommodation segment, is very likely that some solutions may be developed with the support of the national tourism authority, in the scope of the digitization strategy for the sector.

The next steps, where the partners are now working, must include a specific legal framework, both for reporting obligations and permission to share with partners, in the scope of their specific duties and responsibilities. This step is essential to guarantee the legal aspects of SiT.

As for the most challenging aspects of the implementation, the partners identify the heterogeneity of respondents, and how to ensure that the majority have technological and computer literacy skills, as well as the investment to adapt management systems to TAD, with possible "purchase" of a module available in third-party applications, can create resistance.

As a further development, after its implementation, INE intends to coverage to be extended to small Local Accommodation (with less than 10 capacity users), currently not mandatory and thus not covered.