High-quality statistical dissemination:
a strategic goal at the Banco de Portugal

José Faustino[1]

1. Introduction

In recent years, there has been an increase in the number of data requests received at the Banco de Portugal, as well as in the detail and complexity requested by the most frequent and expert users, including international organisations.

This new reality is very much welcomed, but it is also highly demanding. While it is good to know that people are paying attention to the data, at times the data are not being used in the most appropriate ways. All of the effort invested in producing high-quality statistics may be completely wasted, for example, when a newspaper presents the statistics with an incorrect analysis of the figures.

Those who produce statistics play a fundamental role in promoting their use and, particularly, in ensuring that people understand and make the most appropriate and meaningful use of the information.

With the release of a completely new statistical dissemination system, BPstat – Statistics online (http://apl1.bportugal.pt/estatisticasweb//Default.aspx?Lang=en-GB), on 19 January 2006, the Banco de Portugal confirmed that high-quality statistical dissemination is one of its strategic goals.

The rest of the paper is structured as follows. The second section aims at presenting the rationale behind developing a powerful statistical dissemination system. The third describes the data, metadata and functionalities available in the BPstat – Statistics online system. The fourth section focuses on the main components of BPstat – Statistics online. The fifth deals with the importance of disclosing all statistical data to the public. The sixth section is focuses on the inclusion of comprehensive metadata in a statistical dissemination system. The seventh, and last, section presents conclusions and the three main near-term goals for BPstat – Statistics online.

2. High-quality statistics – the main component

The Statistics Department of the Banco de Portugal has, from its inception, been closely involved in producing high-quality statistics. The reporting systems have been designed and implemented in cooperation with data reporters. Most collection systems are based on very elementary data, which reduces the burden on data reporters and allows total flexibility and the adoption of extensive quality control systems, which are a fundamental element in assessing and improving the quality of the statistics.

These collection systems made it possible to build databases with enormous volumes of data, which are a powerful instrument for statistical analysis. Having all of this valuable

---

information and not disclosing it to the public was a less than ideal situation, particularly in view of the increasing number of statistical data requests that have to be analysed and, when approved, answered. Some progress was made with the addition of more tables to the Monthly Statistical Bulletin, but this is not the most appropriate solution when the objective is to significantly increase the volume of information. Instead of the pre-formatted tables, a flexible form of access to data is more appealing to users.

It was essential to adopt a new statistical dissemination policy designed to create a comprehensive and transparent environment in which all non-confidential data and metadata would be available to all users. However, several barriers must be overcome to achieve this. Besides the technical, financial and operational issues, there are other difficulties involved in having open and comprehensive statistical dissemination systems. First, due to a lack of human resources, it is important to implement sophisticated quality control systems when considering the disclosure of high-quality, detailed statistical information. Second, traditional time-series structures would have to be replaced or complemented by multidimensional structures.

In this sophisticated new environment, there are dozens of tasks to be organised, managed and planned. The level of commitment, motivation and technical skills of the system's support staff are the main components needed to assure success.

3. Making data, metadata and functionalities easily available

BPstat – Statistics online is the outcome of a significant investment that the Banco de Portugal has made to provide a statistical dissemination service on the internet. Its main purpose is to provide quick and easy access to the statistics compiled by the Banco de Portugal, as well as to the statistics and economic indicators published by other institutions.

This new statistical dissemination system is completely free. Registration is not mandatory but is recommended for users to benefit from all of the services. The system permits user-friendly navigation through the statistical information and offers several functionalities and services, namely:

- Easy access to timely data (on time-series or multidimensional analysis) and metadata; information updates at least three times a day; and all relevant statistics available (inflation, monetary and financial statistics, as well as statistics on non-monetary financial institutions, balance of payments, international investment positions, public finance, financial accounts, foreign exchange rates, payment systems, banking systems, the Eurosystem, etc), in addition to micro statistics and macro indicators;

- Predefined analysis for quicker access to the main indicators; creation of favourites; a table design facility (static or dynamic analysis, automatically updated when new information becomes available) to suit users' needs, allowing data from different business areas to be included in the same analysis;

- Access to metadata, presented in a clear, logical and standardised manner, including a data publication calendar, a comprehensive glossary, a news service providing information on the most relevant events, and the ability to subscribe to the alerts service that allows users to receive advance information on the publication of new statistical data related to the various domains;

- Fully bilingual information (all information is available in Portuguese and English and deployed simultaneously in both languages); search facilities; graphic capacities; print and export functionalities compatible with major formats (for both data and metadata);
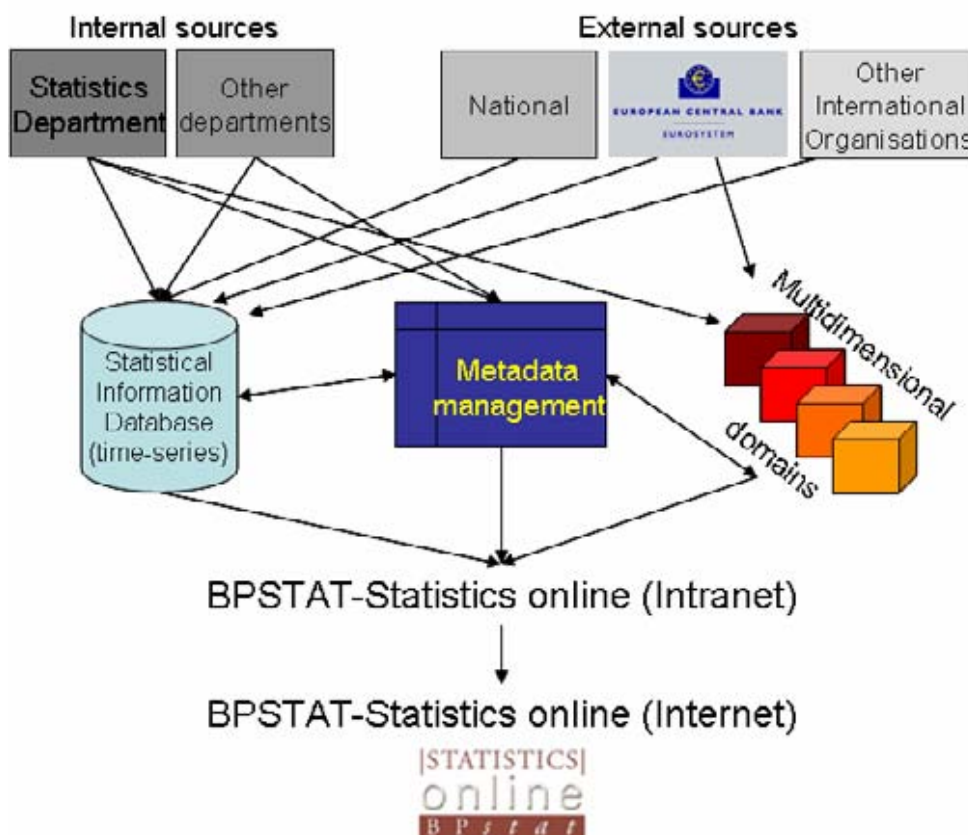
- Help online and through a hot-line during business hours, via various channels – mainly e-mail and telephone; close contact with main users; ongoing investment in recovering historical information and in assessing the quality of the disseminated data; and a record of all instances in which data was accessed and the functionalities used, in order to better understand the demands on the system and how it is being used.

4. Integrated statistical components

Underlying the Banco de Portugal's statistical web service are five integrated and interconnected components. First is a repository of time-series statistics (Statistics Information Database) that includes not only the main statistics produced in-house by all statistical work units, but also those from other institutions that produce statistics. Second, for multidimensional analysis, the cubes are linked to the databases in the statistical production systems, allowing a full and quick "refresh" of data. Third is a metadata management system linked to the automatic update system that refreshes the data and metadata on the intranet. Fourth is the BPstat intranet system, which receives the data and metadata from the three components and performs the final validations. Fifth, after verifying that there are no errors or inconsistencies, data and metadata are copied into the internet version.

Figure 1

The main components of BPstat – Statistics online

Needless to say, the success of a web system depends very much on its performance: the server and communication channels must be prepared to deal with many simultaneous visitors. However, there are associated costs, and it is essential to monitor the system continuously, so as to keep it fully operational, and to determine whether improvements can be made within existing budget constraints.

The increasing number of users overloads the system at certain times, eg on the day the monthly data are released (usually the 15th working day of the month, at 11 am). In order to maintain the credibility of the system, every effort must be taken to ensure that users have no problems accessing the statistics.

5.   Full availability of statistical data

An important goal is to make available all non-confidential statistical data. This requires a huge effort, since it is necessary to first identify all of the statistics that each business area can produce and then assess whether there are any confidentiality issues.

In developing a multidimensional analysis in BPstat, all of the dimensions and elements were identified and a data model for dissemination created. Subsequently, all of the combinations of the different elements of all of the dimensions were assessed, in order to build a "security model" in which all possible combinations that could be disclosed to the general public were identified. In implementing the model, it was necessary to ensure that all rules were properly applied – a requirement that needs to be tracked on an ongoing basis. Once the work is completed and the data are placed in BPstat, there will no longer be any reason to send data requests to the Banco de Portugal, since all data and metadata will be readily available on the internet.

All statistical data produced by the Banco de Portugal are first disclosed through **BPstat – Statistics online**, following a pre-announced timetable, and within a narrow timeframe, to allow users access to the data as soon as possible.

6.   Comprehensive metadata

Presenting metadata cannot consist simply of adding information at the end, as in an old-fashioned hard-copy report, or of having it available in static files, but must instead be conceived differently. For instance, at each level of a statistical domain, including at the observation level, any relevant information can be presented in the form of notes appearing as pop-ups. As stated above, metadata must be presented in a clear, logical and standardised manner.

In 2005 there was a huge investment in metadata by all work units of the Statistics Department at the Banco de Portugal. In 2004, based on a discussion of the best way to include metadata in the new internet-based statistical dissemination system, two main groups of metadata were defined: (1) *reference metadata*; and (2) *support metadata*.

In terms of *reference metadata*, it was decided to keep the approach adopted for the Statistical Information Database. This level of metadata covers all of the characteristics of the series (in the time-series analysis) and the elements and dimensions (in the multidimensional analysis).

Support metadata includes all other metadata, and can be divided into three sub-groups: (a) *context metadata*, consistent with the Data Quality Assessment Framework (see the Special Data Dissemination Standard on www.imf.org), providing information on each statistical domain, using standardised international methods; (b) *simple metadata*, which

include small amounts of text (less than 256 characters) reusable by several components (series, elements, dimensions); and (c) *open metadata*, the only sub-group which follows no predefined structure and covers documents, pictures, tables, etc, and can be made available in the BPstat system or, in most cases, via a link to the website of the institution that has produced them (in most cases the Banco de Portugal, the ECB and the IMF).

Two necessary elements are: (1) a metadata management application that allows each statistics work unit to access all metadata and to manage its own metadata (and also manage "news", user profiles, access, and logs); and (2) an established procedure to validate new metadata in order to ensure that it is consistent with the defined standards.

7.   Conclusions and prospects

While we are confident in the approach we have adopted regarding statistical dissemination, we are also aware of the work that remains to be done.

BPstat – Statistics online takes account of the needs of less skilled, less demanding users, by employing the traditional time-series analysis, while at the same time including the multidimensional analysis which has been very well received among more intensive users of statistics. Our experience shows that most people learn quickly and are no longer looking for predefined tables, preferring instead to construct their own personal statistical dissemination system using the available functionalities.

There is plenty of room for improvement to add value to statistics, help people understand the figures and encourage them to increase their use of statistics.

BPstat – Statistics online has three main goals for the future: (1) to increase coverage (of data and metadata), mainly in the realm of multidimensional analysis; (2) to expand the functionalities, improving accessibility so that users can make use of the data directly or via their own applications; and (3) to contribute significantly to statistical literacy – eg by strengthening contacts with users and by providing more detailed metadata on the concepts, methodology and statistical processes associated with the production and dissemination of statistics.