Reaping the Benefits of Using Integrated Micro-Data for Statistical Purposes… and Beyond

João Cadete de Matos[1] | *Banco de Portugal, Lisbon, Portugal*

Abstract

This paper examines how the *Banco de Portugal* has been exploring the statistical micro-databases available in-house, with the aim to enhance the effectiveness and efficiency of its statistical system.

The use of integrated granular data constitutes the cornerstone of the Bank's long-term strategy as regards its statistical function, and a vital contribution to other areas within central bank's competencies – *inter alia*, monetary policy, financial stability and supervision.

Some examples of areas in which the integrated management of micro-data has added considerable value to the statistics of the *Banco de Portugal*, in particularly by assessing the heterogeneity hidden behind aggregate numbers, are shown.

Keywords	JEL code
Micro-database, data integration, granular data, central bank statistics	*C81, E58*

INTRODUCTION

One of the most significant present characteristics is the constantly changing environment. From medicine to biotechnology, computers to cell phones, there are numerous areas where change has been deep and long-lasting. More than the wide scope of change itself, it refers also to the speed and rhythm with which reality keeps evolving: the changes our sons will witness in the course of their lives will likely be broader than those we are currently observing and even more so compared to those our forefathers experienced.

The financial world is perhaps one of the most affected by the increased pace of innovations. This calls for central banks to be particularly attentive in the fields of financial supervision and regulation, but it also demands from the statistical function to be able to devise solutions that can be quickly adjusted to this developing and challenging landscape, as well as to new and unforeseeable data needs.

---

[1] Statistics Department, Av. Almirante Reis 71, 1150-012 Lisbon. Portugal. E-mail: jcmatos@bportugal.pt, phone: (+351) 213128342.

Against this background, the paper discusses how the *Banco de Portugal* (hereinafter also referred to as "the Bank") has been exploring the statistical potential of a number of available micro-databases, which cover different areas of the economy and the financial system, with the aim to enhance the effectiveness and efficiency of its statistical system while keeping the respondents' burden at an acceptable level. The granular nature of such information, together with a good coverage of the relevant population, offers an increased flexibility as regards the compilation of new statistics and a more rapid response to *ad hoc* data requirements and users' requests.

The use of integrated micro-databases for statistical purposes constitutes the cornerstone of the Bank's long-term strategy as regards its statistical function, and a vital contribution to other areas within central bank's competencies – *inter alia*, monetary policy, financial stability and supervision. We believe that this approach will pave the way to better address the challenges that lie ahead, whichever they may be.

1 WHY MICRO-DATA?

Managing highly detailed and granular databases is the first step of a broader twofold approach; the second is to build a fully integrated data infrastructure. This new integrated management of information approach is expected to generate many benefits (Menezes & D'Aguiar, 2013). On the one hand, it should eliminate unjustified redundancies and lead to the definition of efficient mechanisms for compiling information; on the other hand, it should contribute to the improvement of the data quality and integrity, as well as facilitate the dissemination and agile consultation of the information.

One thing that we have learned during the global financial crisis is that aggregate figures are not sufficient to fully grasp developments in economic variables as they refer to the average of distributions. Quite the contrary, these data should be complemented with micro-data, which enable exploring the heterogeneity hidden behind aggregate numbers. In fact, in many situations, the tails of the distribution provide the most important information, which clearly explains why these data became crucial in recent times.

Accordingly, a move towards micro-data has gradually been advancing at the Bank in recent times, based on and profiting from the many micro-databases managed by the Statistics Department. For instance, the Securities Statistics Integrated System (SSIS) is a security-by-security and investor-by-investor database of both securities holdings and issues. Other such examples are the Central Credit Register (CCR) – which contains granular information on *e.g.* credit exposures – and the Central Balance-Sheet Database (CBSD) – which holds accounting and financial information covering exhaustively the population of non-financial corporations (NFCs).

All of the aforementioned databases provide comprehensive information concerning their respective domains and are extremely rich. However, to reap the maximum potential of these databases, it is essential to take the additional step and, instead of viewing them in isolation as standalone data repositories, linking them in a single fully integrated high granular data system. By linking the information contained in each individual database, this data system will boost the potential associated with each one, enabling the crossing of data on different institutional sectors and financial instruments.

1.1. Clear benefits for statistical production...

This fully fledged integrated system, encompassing granular data of all institutional sectors and financial instruments, serves the purposes of the different statistical domains, which can, in turn, feed the system with the information they produce, while at the same time tapping into the system for the information they need. In particular, to the extent that they put together all sectors of the economy in a single framework, in an integrated and balanced manner, national financial accounts stand to benefit significantly from such a data system.

Figure 1 schematically illustrates this point. It displays a matrix with institutional sectors in column and financial instruments in row. The dimensions that are currently covered by the Bank's

micro-databases are highlighted in green, while those that are deemed feasible in the short/medium-run are highlighted in yellow. In more detail, the SSIS gives us granular information on all kinds of securities; the CCR has micro-data regarding loans to all sectors; the CBSD gives us a complete view on the non-financial sector assets and liabilities; the Balance Sheet Information (BSI) on Financial Corporations has granular information on the assets and liabilities of the sector; the BoP/IIP system supplies micro-data on the assets and liabilities of the Rest-of-the-World sector. To complete the few gaps in this matrix, it would be feasible to get information on "Insurance and Technical Reserves" and also more granular information regarding the "General Government" sector. However, our most immediate goal is to get granular data on "Currency and Deposits", which would allow us, in the shortest possible time span, to have very reasonable micro-data coverage of the economies' total financial assets and liabilities.

To sum up, a significant amount of this endeavour is well underway and only a few steps – some of which require legal support – are needed to achieve full completion, mostly concerning the household sector.

Figure 1 National financial accounts – Assets and liabilities, by institutional sector and instrument

Note: NFC – Non-financial corporations; FC – Financial corporations; GG – General governement; HH – Households; NPISH – Non-profit institutions serving households; RoW – Rest of the world; SSIS – Securities Statistics Integrated System; CCR – Central Credit Register; CBSD – Central Balance Sheet Database; BSI – Balance sheet information; BoP/IIP – Balance of payments and international investment position.
Source: Banco de Portugal

1.2 ... but also for users and for analytical purposes

Providing more complete and detailed statistics in response to users' needs is a fundamental objective to be pursued by the statistical function. This became particularly evident, namely with the eruption of the global financial crisis, which sparked a whole array of new data needs.

The financial turmoil of 2007–2009 highlighted potential (and actual) gaps in the statistical framework, both at national and international level. In particular, the crisis revealed important gaps in information for the purposes of financial stability analysis, namely concerning counterpart data. The development of micro-databases and administrative records reporting can show a major contribution in overcoming some of these shortcomings (Lavrador, 2010). They permit to develop knowledge about the activities of economic and financial agents at a more detailed level and allow for the drawing of conclusions that would not be possible should one rely solely on aggregated data.

The international financial crises also emphasised the importance of timely, efficient and reliable financial data to support monetary policy and financial stability decisions. Furthermore, the new demands and requirements for the compilation of statistics also revealed that most of the traditional

statistical methods of compilation and analysis do not meet all the data quality requirements. Since then, the development of new methods and tools to manage and improve data quality in financial statistics has become a reality in Statistical Units in many countries. These tools support organisation-wide analysis, the integration of different statistical domains, and enable complex analysis to be made available to decision making bodies (Aguiar & Lavrador, 2012).

One illustrative example where the features of micro-data have proved to be very useful can be found in the analysis of the non-financial sectors indebtedness. Using the available micro-databases, namely the CBSD, the SSIS and the CCR, and taking advantage of the reference tables and related administrative sources, the Bank started to publish, in February 2012, very detailed statistics on non-financial sector indebtedness, with several different breakdowns. First of all, information on the counterparty financing sectors is provided, so as to ascertain the risk exposures of creditor sectors, as well as possible funding strains of debtor sectors. Second, credit to NFCs is broken down according to maturity, as well the economic activity sector and enterprise size; along the same lines, household sector loans are also disaggregated by purpose (housing loans versus consumption and other purposes). Finally, special attention is paid to the public sector. For instance, different public debt definitions are provided – consolidated, non-consolidated, Maastricht debt (relevant for Excessive Deficit Procedures), including/excluding state-owned enterprises.

Another example concerns the usefulness of merging accounting information at the company level from the CBSD with data from the CCR, as a means to analyze the drivers of company's credit risk. This line of research allows for identifying the emerging risks in banks' portfolios, as well as creating modelling tools for the forecasting of default probabilities. In fact, the Bank has recently taken decisive steps towards further exploring the informational potential of the CCR and balance sheet databases in an ongoing project that aims at creating an internal credit assessment system (ICAS). This system will provide the Bank with its own in-house structure to assess credit risk, thus reducing its dependence on external sources. Against the background of the global financial crisis and the shortage of assets liable to be used as collateral in monetary policy operations, these systems have recently been gaining importance within the Eurosystem, as can be seen by the increasing number of NCBs that have introduced them or are planning to do so, namely Austria, Belgium, Germany and Spain.

More broadly, and to sum up, micro-data applications have several other uses in many different fields of central banking. Besides being vital for economic and financial research, they have been progressively gaining a more relevant role in other areas, such as financial stability and supervision activities, monetary policy and risk assessment.

2 HOW TO INTEGRATE MICRO-DATABASES?

Needless to say that achieving such architecture is not an easy task, one which cannot be accomplished overnight. Attaining these goals hinges on an effective cooperation between different functions of the Bank, based on sharing of knowledge and the identification of the information needs of both users and producers. A stepwise approach is warranted, whereby the integration proceeds in a gradual and phased manner. In what follows, I will go through these aspects in more detail.

2.1 Governance structure

The definition of an Information Governance Structure aims to ensure a proper alignment between the strategic and operational levels of decision, which are mediated by the integrated management of information (see Figure 2).

Figure 2 The governance structure
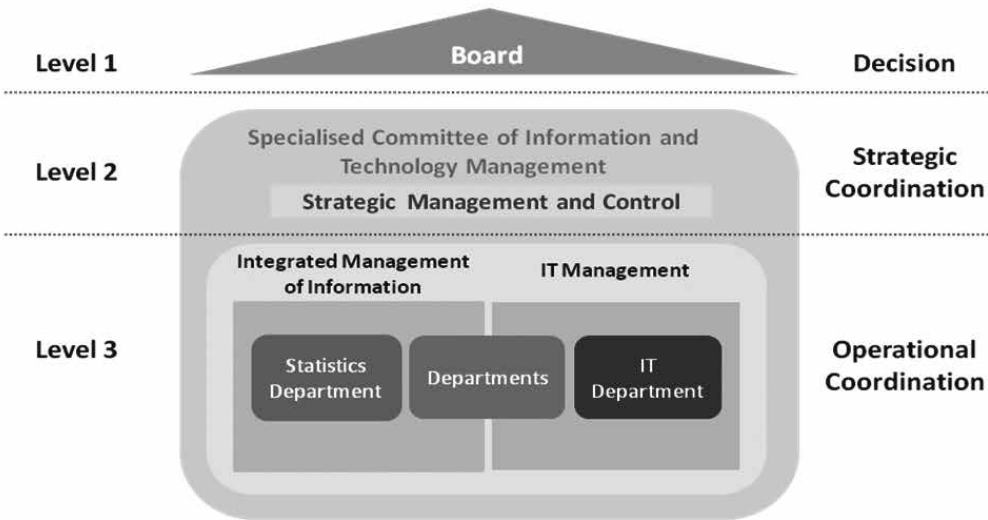
Source: Banco de Portugal

The Statistics Department is in charge of the operational management, including:

a. Coordinating and monitoring the process of collecting quantitative information from external entities;
b. Ensuring the central point of contact of the Bank with external entities on the reporting of quantitative information;
c. Promoting, in conjunction with the IT Department and the user departments, the:
– Organization of information architectures, namely by identifying objects, features and respective relationships and configuring the domains of integration to manage.
– Definition of concepts and creation of metadata associated with different information objects in order to avoid duplication and facilitate the understanding/utilization of information.
– Creation of catalogues/dictionaries/repositories of information available on particular operating systems.
d. Monitoring the interaction and timely reporting of information to and from external entities.
e. Analyzing the changing needs of quantitative information identified by other departments.
f. Guaranteeing the quality of information, defining indicators of their use and ensuring its relevance and auditability.

In this context, various departments that are originators/users of information have the decentralized responsibility, in collaboration with the department accountable for the centralized management of information, of analyzing in a critical manner the data and the metadata that are most important to them and ensure their quality. They also collaborate on the identification of the functional requirements, having in mind the integrated and shared management of information – the identification of functional requirements is the basis for the consolidation of logical and technological architecture.

2.2 Relationships management

Given the large number of stakeholders, a relationships management is essential, namely to introduce greater efficiency in the communication process, normalizing and formatting it in the customer's perspective. It is based on two cornerstone principles:

1.   Information is a key asset of the Bank so it must be managed in an integrated way.
2.   The exploration and analysis of information are distributed activities, typically related to the needs and tasks of each department.

Moreover, an efficient management of information should be based on shared management, which requires a separation of responsibility between the "originator/user of information" and the "manager of information". The first is best done in a decentralized way by each department, while the latter should be concentrated in a single department. In fact, given that information is a common good, it should be managed by specialists – these specialists are better placed to collect, classify, manipulate, store, recover and disseminate information.

2.3 Information architecture

The infrastructural base of the information management – the so-called Information Architecture – should also be mentioned in this context. Its main aim consists of ensuring the quality, auditability and manageability of the data. It also serves to establish levels of responsibility in the management of information, separating the activities related to the organization and processing of information from the analysis and exploration activities. It is based on five layers where the division between the information management and the exploration and analytics activities occur from the 3rd to the 4th layer, as can be seen in Figure 3. To successfully integrate the different domains, reduce the reporting burden and eliminate redundancies at several levels, it is important to develop high quality reference tables and to maintain up-to-date metadata and catalogues.

Figure 3 Information model
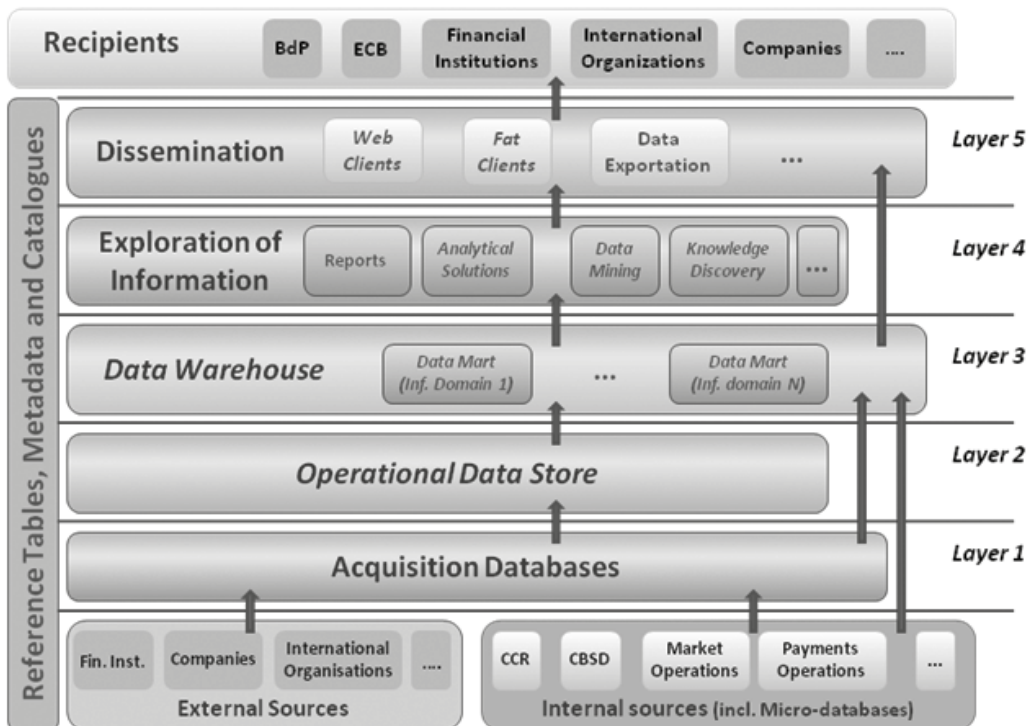
Source: Banco de Portugal

CONCLUSION

The Statistics Department has been pursuing a strategy centred in the integrated management of highly granular data. We believe that this approach will provide us with necessary tools to answer both aggregated and highly detailed queries. Moreover, crossing the different data dimensions will also lead to higher quality standards and efficiency, minimizing the reporting burden. Furthermore, we believe this is extremely useful for our users, both internal and external, to the extent that it enables us to provide tailor-made data with a shorter reaction time.

We cannot anticipate what the future will bring. But the good news is that, to a significant extent, we do not need to consult oracles: we can prepare ourselves to better respond to whatever comes along without necessarily knowing what we will be facing. To this end, we need to proactively invest in becoming flexible. In other words, we can endow ourselves with systems and tools with a sufficient level of manoeuvrability that will enable us to adapt swiftly to whatever comes along.

References

AGUIAR, M. do C., LAVRADOR, I. *Data quality management in statistical systems: the importance of business intelligence.* European Conference on Quality in Official Statistics, Athens, May 2012.

ALMEIDA, A. M. de, D'AGUIAR, L., CASIMIRO, P. *Promoting enhanced responsiveness to users' data needs: the experience of the Banco de Portugal in exploring the statistical potential of micro-databases.* 58th World Statistics Congress, Dublin, August 2011.

LAVRADOR, I. *Exploring the statistical potential of micro-databases.* Irving Fisher Committee Conference – Initiatives to address data gaps revealed by the financial crisis, Basel, August 2010.

MARTINS, C. et al. *Arquitectura de BI – Sistema de Informação Estatística.* Banco de Portugal, June 2008 (not available in English).

MATOS, J. C., MARQUES, C. *Innovative solutions to compile balance of payments statistics minimising costs and ensuring quality.* Seminar on Business Related Statistics, Luxembourg, February 2012.

MATOS, J. C. *Drawn to excellence – a sample of issues critical for compiling and disseminating central bank statistics.* Joint Meeting of the y-BIS and the jSPE, Almada, July 2012.

MATOS, J. C. *Innovative solutions in compiling financial accounts.* Workshop on Developing and Improving Sectoral Financial Accounts, IFC and Central Bank of the Republic of Turkey, Istanbul, May 2014.

MATOS, J. C. *The information model at Banco de Portugal: innovative and flexible data solutions.* Meeting on Financial Information Needs for Statistics, Macroprudential Regulation and Supervision in Central Banks of Latin America and the Caribbean, CEMLA and Banco de México, Ciudad de México, May 2014.

MENEZES, P., D'AGUIAR, L. *Reaping the benefits of integrating the micro-databases available at the Banco de Portugal.* XIX Jornadas de Classificação e Análise de Dados (JOCLAD 2012), Tomar, March 2012.

MENEZES, P., D'AGUIAR, L. *Impact and Benefits of Micro-Databases' Integration on the Statistics of the Banco de Portugal.* 59th World Statistics Congress, Hong-Kong, August 2013.

New chapter on non-financial sector indebtedness. Press Release, Banco de Portugal, February 2012.

SANCHAS, R., AGUIAR, M. *Business intelligence in statistical systems: a stepwise approach.* XIX Jornadas de Classificação e Análise de Dados (JOCLAD 2012), Tomar, March 2012.

Supplement to the Banco de Portugal Statistical Bulletin 3/2013. Papers presented in the Workshop on Integrated Management, Porto, June 2013.