Business intelligence in securities statistics: an evolutionary approach

Maria do Carmo Aguiar[1]

"It is not the strongest or the most intelligent who will survive but those who can best manage change."

Charles Darwin

Storyboard

Securities statistics at Banco de Portugal (BdP) have been around for a while. In fact, they were first released in the BdP's Annual Reports of the 1940s and their regular publication, with information on the issuance of short-term securities (e.g. Treasury bills and commercial paper) and long-term securities (shares and bonds), dates back to the 1980s. Monthly data on net issues and end-of-period positions, including totals and values broken down by type of security, maturity, currency, institutional sector of the issuer and respective classification of economic activity, were released for the first time in January 2005.

A milestone in the history of BdP securities statistics occurred in the late 1990s with the creation of the Securities Statistics Integrated System (SSIS). This system was developed with the purpose of gathering, in a single repository, all the information deemed necessary to comply with reporting requirements on securities.

The SSIS stores securities issues and portfolios (holdings) data on "security-by-security" and "investor-by-investor" basis. It comprises detailed data on stocks and transactions of debt securities, shares and other equity (financial derivatives are not included). Both stocks and transactions are collected on a monthly basis. The classification of securities relies on the existence of a reference database which, in the case of domestic securities, is maintained by BdP; foreign securities are classified according to the Centralised Securities Database (CSDB) managed by the European System of Central Banks (ESCB).

In what respects to securities holdings, detailed information is collected on the investments by residents in domestic and foreign securities, as well as on the holdings of non-resident investors in domestic securities. This information is reported by financial institutions, both as investors and custodians, and other resident entities.

SSIS promotes consistency across statistics produced by BdP and makes it possible to meet user needs, both at national and international level. In fact, SSIS information is used as an input for the compilation of a wide set of statistics produced at BdP, from monetary and financial statistics to balance of payments and international investment position and financial accounts. On the other hand,

---

information from other statistical sources is used as an input for estimation of missing data and for data quality management purposes.

The initial SSIS architecture relies on two relational databases: the collected data is stored and validated in the "transactional database" (which is also used to estimate missing information); validated and "enriched" data are copied to the "exploration database" on a daily basis. In 2006, the analytical capabilities of the system were enhanced through the development of an analytical database, which enables user friendly multidimensional analysis of the information.

While multidimensional analysis of securities data has proved to be a powerful tool for rapid analysis and reporting, we foresee advantages in adding dimensions of analysis from other statistical systems, whose integration is not straightforward. Thus, more than ten years after its inception, it was time to reformulate SSIS. Launched in 2012, an Information Technology (IT) project aims at developing a new SSIS within the Business Intelligence (BI) framework envisaged for the BdP statistical systems. The following section describes the main features of this framework.

Behind the scenes

Evolution of methodologies, frameworks and statistical demands are the main reasons for the need to reformulate most of the information systems that are used in statistical production. Furthermore, the need to interconnect different statistical domains and cross check data has become a fundamental requisite.

Business intelligence (BI) is "a broad category of applications and technologies for gathering, storing, analyzing, and providing access to data to help enterprise users make better business decisions". Simply put, BI is about getting the right information to the right people at the right time, so that they can make good decisions that improve organisational performance.

In the beginning of 2008, the IT Department and the Statistics Department of BdP initiated a study with the purpose of defining a BI framework, to be used as a reference in all future IT developments in the statistical area. This framework is built upon three pillars: a data warehouse, centralised reference tables and a common IT platform. The data warehouse guarantees a central access point to every statistical data, independent of the input source or the production process; the centralised reference database provides common reference data and enables cross linking information from different sources and systems; the consistent usage of a common technological infrastructure across the multiple information systems makes it easier to integrate and reuse components and promotes data access efficiency and transparency to final users.

The stocktaking exercise on the relationships between existing statistical systems showed a high level of interdependency, with multiple sets of data being exchanged over the network, each of which poses a need for data extraction, cleaning, transformation and integration. A data warehouse was proposed to deal with this overhead of multiple data exchanges. Each new or renewed system would contribute to, and use data from, a single centrally managed database. The goal was to achieve a situation where there is a single version of trusted data, and little or no data is directly exchanged between statistical systems.

The second pillar of the BI framework is the Master Data Management (MDM), which collects, consolidates, stores and delivers reference data (e.g. countries, currencies, financial sectors,...) that are used across the systems. A register of financial institutions has been kept for long, and there was an effort, still ongoing, to streamline the process of gathering data from several sources and consolidating it into an historical register of all resident companies. The MDM assigns each register an internal surrogate key that never changes; the mapping between the internal key and the corresponding external business or natural keys (different data sources often report different keys for the same data entity) are kept in the MDM system. This is in fact an extremely important component of the framework, as it gives consistent meaning to the facts stored in the data warehouse, and makes much easier the crossing of fact data from different systems.

In what respects to the third pillar, the IT platform, in the last quarter of 2007, the IT Department and the Statistics Department jointly carried out a Proof of Concept (PoC) for the selection of the BI Platform to be used in the forthcoming IT projects of the statistics business area. After a preliminary market analysis, and given the history of BI tools usage at BdP in the recent years, the short list was reduced to two solution providers: Microsoft and SAS. The results of this PoC showed that none of the two platforms could deliver a high level of functionality across the whole spectrum of business needs, but they were in fact complementary. The final choice was then a combination of the two solutions.

Meanwhile, BI tools and concepts have evolved and the focus nowadays is much more on the so called data discovery tools, which enable more powerful analysis of data from multiple sources, which do not necessarily imply a centralised data warehouse. Against this background, a new PoC is currently taking place at BdP in order to re-evaluate the best BI platform solution to accomplish our targets.

To be continued...

In a nutshell, statistical systems comprise four phases: Acquisition, Processing, Exploration and Dissemination. Each phase has a comprehensive set of core tasks and processes. Redesigning the SSIS according to the BI framework implies redefining these processes, along with adopting the common IT infrastructure. The data model that supports SSIS needs to be redesigned in order to separate data acquisition and data processing repositories, which currently are the same. Internal reference data will be replaced with MDM references, ensuring harmonisation with other statistical systems. When available, input from other systems will be provided via the data warehouse; this will be an incremental process. Multidimensional analysis, already being performed, will benefit from data of other systems, available via the data warehouse. Consequently, efficiency gains are expected, along with improved data analysis.

BI researchers state that "organisations which have the greatest success with BI travel an evolutionary path, starting with basic data and analytical tools and transitioning to increasingly more sophisticated capabilities". On the other hand, market leaders in BI are constantly providing new tools to address the main issues in this field: integration, information delivery and analysis.

In statistics, like in many other areas, there is the need for continuous improvement and innovation. A stepwise approach is not only wise but the most realistic to be followed.

References

Aguiar, M (2007). "Integrated Statistical Systems: evolution or revolution?", 56th ISI Session (Irving Fisher Committee, IFC Bulletin N. 28), Lisbon, August 2007.

Aguiar, M & Martins, C (2011). "Adding Business Intelligence to statistical systems – the experience of *Banco de Portugal*", Eurostat Conference on "New Techniques and Technologies for Statistics". Brussels, February 2011.

IBM & Gartner (2011). "A step-by-step approach to successful Business Intelligence".

Nelson, Greg (2007). "Introduction to the SAS® 9 Business Intelligence Platform: A Tutorial", SAS Global Forum 2007.