

## Guidelines for a flexible and resilient statistical system: the architecture of the new Portuguese BOP/IIP system

Marques, Carla

*Banco de Portugal*<sup>1</sup>, Statistics Department

Av. D. João II, Lote 1.12.02

Lisbon (1990-204), Portugal

*csmarques@bportugal.pt*

Currently statistical compilers face a challenge of implementing efficient statistical systems that can adapt to constantly changing reality, not only in terms of data requirements and outputs but also in terms of data sources. In this context it is of utmost importance to build flexible and resilient systems. One other matter of concern is the adequate balance between costs (for the compilers and most importantly for the reporting entities) and quality measured as the relevance and accuracy of the data provided to users.

This paper presents the Portuguese experience in designing a new system for the collection and compilation of balance of payments (b.o.p.) and international investment position (i.i.p.) statistics which will replace the current settlement-based system established in 1993. It is based on a business intelligence general architecture designed as a benchmark to be followed by new statistical systems in *Banco de Portugal* which is built upon three pillars: a data warehouse; a centralized reference tables and; a common IT platform.

The new b.o.p./i.i.p. system is described along four major phases (acquisition, processing, exploration and disclosure) highlighting in each phase its guiding principles: centralization; harmonization; flexibility; consistency and efficiency.

### 1. The challenges and the drive behind the change

The current collection and compilation system for balance of payments (b.o.p.) statistics was set up in 1993 upon the liberalization of the capital movements in Portugal. This system was mostly based upon settlements' data reported by resident banks, which reported transactions with non-residents on their own behalf and on behalf of customers, with all the necessary statistical classification. Furthermore, all entities were obliged to report transactions with non-residents settled without the intermediation of the resident banking sector. This would be the case of transactions settled through an account abroad. The system was completed with other specific data sources such as *Banco de Portugal* (BdP) own transactions and portfolio investment surveys.

As of late nineties the system developed to incorporate new data sources and to meet new statistical requirements. National Statistical Institute information on external trade in goods, direct investment annual surveys, securities statistics integrated system and money and banking statistics are some of the additional data sources incorporated during this period. On the other hand, new statistical requirements include the compilation of international investment position (i.i.p.) statistics (as of 1999 with back data as of 1996) and other requirements, namely those associated with the Portuguese participation in the European Community.

The first motivation for changing the current collection and compilation system is thus to move from a multiple heterogeneous system that resulted from the various developments described previously, to an integrated and coordinated system that takes on board the current complexity in compiling b.o.p. and i.i.p.

---

<sup>1</sup> The analyses, opinions and findings of this paper represent the views of the authors; they are not necessarily those of *Banco de Portugal* or the Eurosystem.

statistics. Namely the existence of distinct and heterogeneous data sources, with different periodicities and granularity, and new and more demanding statistical requirements. The challenge is to do it without overburdening the reporters or the compilers, while insuring a coherent final data set with good quality.

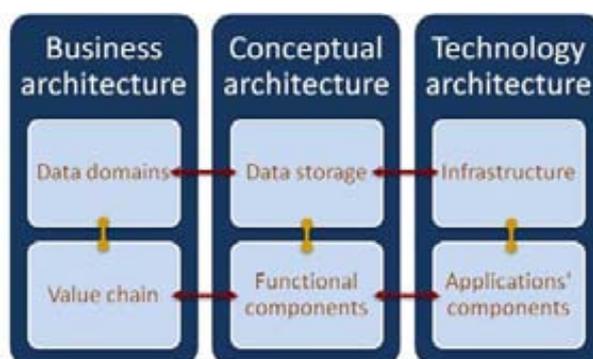
The need for change is emphasized by the current limitations in the use of banks' settlement data. Firstly, globalization of international markets and integration of payments systems, together with increasing significance of multinational groups that resort to cash pooling, clearing practices or treasury centers, amplified the volume of b.o.p. transactions performed outside the banks' payments system. Secondly, the European banking community pressured for the introduction of restrictions in the statistical reporting. The new Regulation (EC) n° 924/2009 on cross-border payments in the Community establishes a €50.000 threshold for settlement-based national reporting obligations on banks and other payment service providers, for balance of payments statistics, for transactions on behalf of their clients. Additionally it is under review the possibility of establishing a total exemption of such obligations for transactions in euro within the Community. The challenge is to comply with these limitations while maintaining the same level of quality statistics.

Finally Bdp statistics comprise heterogeneous but interdependent statistical systems, including macro and micro databases. Currently Bdp manages three item-by-item databases: the Securities Statistics Integrated System (SSIS), the Central Credit Register (CCR) and the Central Balance Sheet Database (CBSDB). The ability to provide an efficient and reliable response to statistical requirements and analysis relies not only upon data availability but also in the ability to perform quick multidimensional analysis correlating different data sources and statistical domains. Within this framework Bdp established a goal of building an integrated and coordinated information system for all statistics produced by the Statistics Department. This long term goal is to be achieved by the definition of a Business intelligence (BI) architecture benchmark to be followed by new statistical systems in Bdp. Within this framework the last reason to change the system is to adopt de BI architecture and to benefit from the latest IT tools.

## 2. Guidelines for the new statistical system

### 2.1. Business intelligence (BI) architecture

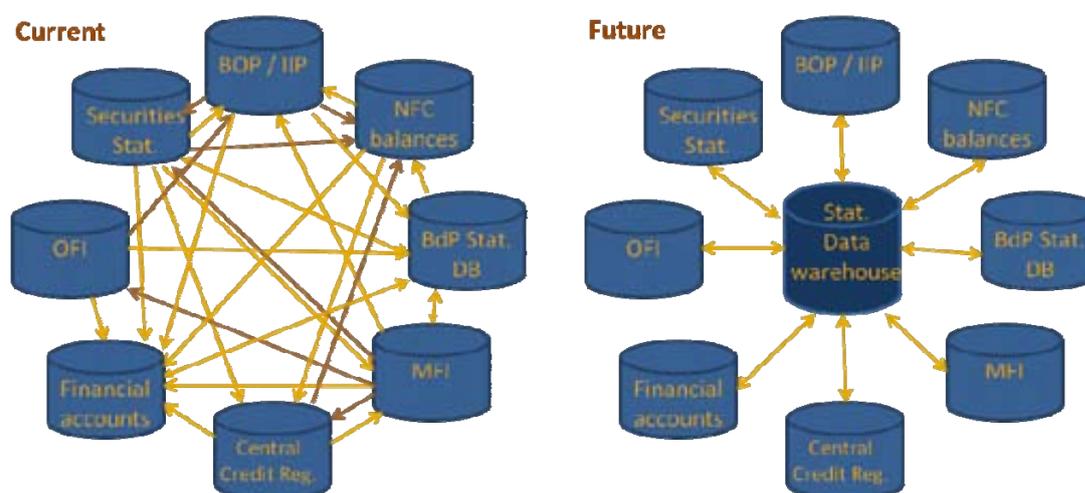
The BI architecture was developed following a three layered approach to the problem: the business layer; the conceptual layer; and, the information technology layer. For each one the structural and functional features of the problem were tackled and analyzed, combined to produce an articulated and coherent perspective of the BI architecture.



The Statistical Department BI architecture benchmark is built upon three pillars: a data warehouse; centralized reference tables and; a common IT platform.

**A statistical data warehouse (SDW).** The various statistical systems within BdP are interconnected, not only at the compilation level as the data produced by a system is used as input data for another system, but also at quality control level in the sense that cross checking data of different statistical domains is essential to ensure a suitable articulation between different statistical outputs and enhance the overall quality of statistics. Each relation implies the development and maintenance of a flow of data that, so far, were ensured in an individual case-by-case basis, following specific approaches. This solution proved to be inefficient and heavier to maintain in the long run, and clearly justifies the development of a data warehouse, which would provide a unique and harmonized repository of data (as detailed as needed).

Each new or revised system will feed its data into the data warehouse and use the data already available in it. This is a gradual process to reach a situation where virtually every data exchange between different statistical systems is established through the single centrally managed database. This will insure a central access point to every statistical data, regardless of the input statistical process, with clear productivity gains in the developing and maintaining of the various data flows. The following figure illustrates the current and future situations.



**Centralized reference tables.** The second pillar of the BI architecture benchmark is the existence of centralized reference tables, which provide a common ground in terms of reference data and facilitate the interlinking between different statistical systems.

This scheme is already in place within BdP for several years. A technical team identifies reference data of interest for several domains (statistical, supervision, accounting or other); defines the owner of the data and provides the tools for updating and managing it; ensures the storing and maintenance of the data; and, supplies the data to all the users throughout the BdP. Some of the reference tables already considered within this scheme refer to countries, currencies, institutional sectors, maturity breakdowns and financial entities. Currently efforts are being made to establish a common reference table for non-financial entities, covering structural and quantitative features (the former include information like name, fiscal number, residency, economic activity classification and institutional sector classification whereas the latter refers to other variable features like capital and number of employed people).

**A common IT platform.** The consistent usage of a common technological infrastructure across the statistical systems facilitates the integration and re-usage of components, promotes data access efficiency and communication between statistical systems. The IT Department carried out a preliminary analysis and sanctioned a solution based on Microsoft (SQL Server 2008) for the structural components and SAS for functionalities in the domain of statistical analysis and analytical workflow. Microsoft has also been

indicated for other utilities such as multidimensional models, ad hoc analysis and dashboards.

## 2.2. Other guiding principles

**The maximization approach.** One other guideline valued in the development of the new system is the maximization approach. Regardless of the statistical reporting requirements in terms of output, and the currently available source data, the entire b.o.p. and i.i.p. statistics will be compiled with the maximum detail and on a monthly basis. The database that supports the compilation of these statistics will have a common structure that comprises all the relevant dimensions. This approach better guarantees the coherence between different breakdowns and across periodicities (monthly, quarterly and annual outputs), it also allows to respond to new requirements without the need to change the basic databases structures, or to include new data sources. Obviously the quality of the statistics will depend on the available data sources, but the decision in terms of what to disseminate, and with what detail and periodicity, will be made when crossing to the dissemination phase, and not in the processing phase.

Additionally, whenever possible, the data will be obtained on a micro level, operation by operation, item by item. Although this option requires dealing and managing a large volume of data, it also grants higher compilation flexibility with lower reporting costs.

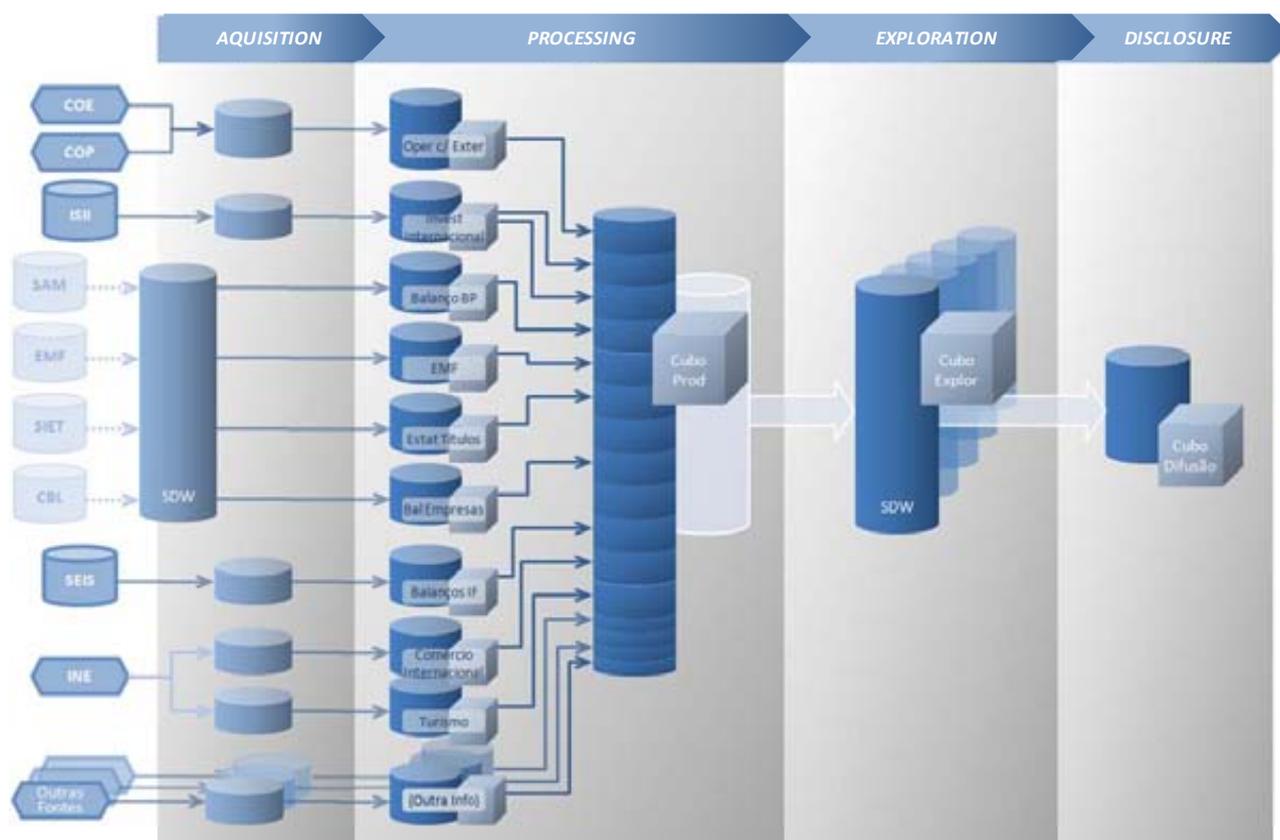
**The powered user approach.** It is an essential condition for a resilient system that the user has the ability to adapt the various processes according to different methodologies or ad-hoc requests, without the need for IT development. This guideline reflects in various stages of the system, namely: the tests and processes are associated to reference tables that can be edited by qualified users; the core data structure is implemented in MS SQL Server, but whenever possible or necessary, the processes are built in SAS; virtually every analysis can be developed or modified by the user.

**Filling the gaps.** Each item was scrutinized individually, considering the available data sources, its strengths and weaknesses. There was an effort to avoid redundancies and to exploit already available databases: if a given information was already reported to BdP or otherwise available within an accessible and reliable data source, then that information was not considered in terms of the obligation to be imposed on the reporting entities. There was also a concern to identify the entities in the best position to provide the necessary information. This exercise was of particular importance to avoid imposing an unnecessary burden to the reporters and to ensure the quality of the statistics.

## 3. The new system

The b.o.p. / i.i.p. statistics system is quite complex and challenging. On one hand, there is not a one-to-one relationship with the data sources: each item will be compiled on the basis of a different set of data sources that may, in some cases, differ according the institutional sector. On the other hand, it is essential that the output is coherent and articulated. The b.o.p. is a double entry system but in the actual compilation process each 'leg' of the operation might be reported by different data sources. It is thus a challenge to guarantee that the whole operation is recorded in a coherent fashion and in the correct items.

The new system will be briefly presented following the value chain perspective within the business architecture: Acquisition; Processing; Exploration; and, Disclosure. Each phase has a set of different tasks associated, different data storages, and different users, but was developed following some common guiding principles: centralization, harmonization, flexibility, consistency and efficiency.



### 3.1. Acquisition

The acquisition constitutes the first phase of collecting input data from external and internal sources. Other tasks performed in this phase are the validation of the file format and the respective source, data loading and a first automatic data quality check.

There are several external data sources, with different features and complexity, with different periodicities and different natures (statistical, accounting, settlements, ...) so it is essential that the acquiring process is flexible enough to adapt to each specific source: there is no one-format-fit-all.

The internal sources comprise data from the statistical department but also data provided by the accounting department (essential for the compilation of the Monetary Authority and Reserves assets statistics) and other departments within BdP. The data from the statistical departments comprises not only statistical data (monetary, financial non-monetary and securities) but also data from administrative databases managed by the department (central balance sheet database and central credit register). Regardless of its nature, the process is very straightforward when the data is stored in the SDW, as data is already validated and stored with the correct (common) codes.

The b.o.p. external source can be organized in the following categories:

- Direct report of external operations
- Settlements data, immediately available within resident banks
- External data from other statistical entities and organizations (including the National Statistical Institute data on external trade of goods)
- Specific surveys targeted to fill in particular and limited gaps

Direct report of external operations is one of the main data source, and one of the most challenging. The shift from a settlement based system to direct reporting by the entities that perform the operations implies an increased burden for these entities, especially for non-financial corporations that were not as

familiarized or aware of the statistical requirements. The challenge is to handle an extraordinary increase of the number of reporters and to obtain their cooperation in providing good quality data within the established deadlines and according to the required format. This challenge was tackled in a three action plan: first, to define statistical requirements according to the perspective of the reporting entities. This requires an additional effort from the compilers that, within the acquisition phase, have to convert and standardize the data according to the b.o.p. and i.i.p. perspective. Secondly, to develop several auxiliary tools to support the reporting, namely by using the banks' settlement data. A reporting entity has three ways to submit the data: to upload a xml file generated internally by its IT system; to access the online application available in BdP website and manually introduce the required data; or, through this same application, to access the payments data provided by banks referring to operations performed to their own entity, and to complement it with the required statistical information (like for example the purpose of the transaction). Thirdly, to feedback useful maps and reports to the reporter, calculated on the basis of available data within the BdP. This third action is essential to change the viewpoint of the non-financial corporations, that consider the statistical obligations as just a cost and neglect to consider the knowledge and information one can obtain through the statistical data.

All the interaction with the direct non-monetary reporters is facilitated and enhanced by the creation of a web page within BdP external site dedicated to these entities. Within this web page non-financial corporations can submit the data (regardless of which of the three methods was chosen), have on-line access to some applications but also receive feedback on the quality of the file submitted and on the data itself, as well as an easy access to useful information (like guidelines and Instructions). This is intended to be a restricted area, with user id and password validation, where the reporting entity will have easy access not only to statistical applications and tools but also confidential information concerning the entity itself.

This restricted web page represents a first step in establishing a new relationship between BdP and non-financial providers of information, and confirms not only a repositioning of BdP but also the growing importance of these institutions.

### 3.2. Processing

The processing phase is where b.o.p., i.i.p. and other statistics are compiled, and is the core of the system. It comprises two sub-phases: a pre-production phase where data quality control, analysis, editing and managing is performed taking in consideration the characteristics and specificities of each data source; and the production phase where all the data is transformed to fit a unique common format, considered suitable for the compilation of statistics.

**The pre-production phase** is specially focused on quality control: internal coherence of each data source; cross validation between different data sources; outlier detection. Data editing resulting from value correction, data estimation and annotation is equally important within this phase. In the last stage, it is decided which data is kept on to the production phase. These data is transformed and standardized to a common statistical format.

**In the production phase** the quality control is performed on a more aggregated level, a common database comprises not only the actually reported data but also calculated and derived measures and estimated data. The core data structure is implemented in MS SQL Server, but most processes are built in SAS, making use of its specialization towards statistical operations, and also allowing for these processes to be adjusted by power users, whenever a different methodology must be implemented.

The production database stores all data relating to the current processing cycle, and although presenting a common structure, it maintains memory of the origin of the data. In fact, for the same item there might be various data sources that are complementary or concurrent, so, for each specific item, data sources they are classified as primary data, supplementary data and cross check data. The actual item will be

calculated by applying a predefined combination of the available data.

It is important to stress that both b.o.p. and i.i.p. are compiled within the same database, guaranteeing an articulation between them. For the same reason, although i.i.p. is only disseminated quarterly, there will be a monthly simulation of this statistics.

The final stage of this phase will be the decision in terms of the data to release, internally to other statistical systems and externally, based on the quality and relevance of the data.

### **3.3. Exploration**

The exploration phase is centered in using the data previously compiled and checked by b.o.p./i.i.p. compilers. It mainly comprises exploration management tasks such as data delivery and reporting, dimensional analysis and time series analysis. The exploration database of the b.o.p./i.i.p. statistics are included in the SDW and have to comply with the rules and conventions that have been determined. The data will be accessed by different type of users within BdP, from other statistical units to economists of the studies department. Each user will have different levels of access to the data but, as a minimum, the data exported to this database will have to have the detail necessary for the most demanding user.

### **3.4. Disclosure**

The disclosure processes are focused on the statistical data dissemination obligations to external entities and also for the general public, it comprises tasks related to statistical publications, statistical reports distribution and output to external databases, namely BDIE (statistical time series database of BdP) and BPSat (a multidimensional statistical dissemination system). The disclosure database presents a lower level of detail and is analysed according to confidentiality and disclosure practices of the BdP.

## **4. Conclusion**

The b.o.p. and i.i.p. statistical system is a very complex one, dealing with multiple data sources and different reporting requirements and obligations. Currently, due to several factors, this system has to be revised giving the opportunity to embrace a new business intelligence architecture and follow other guidelines to ensure a coherent, integrated, flexible and resilient systems. It was also considered the adequate balance between costs (for the compilers and most importantly for the reporting entities) and quality measured as the relevance and accuracy of the data provided to users was also considered.

The new system is briefly described along four phases: Acquisition; Processing; Exploration; and, Disclosure, presenting the main points in each phase. A special emphasis is given in terms of the acquisition of data from direct reporters, one of the main data sources and one of the most challenging, with the development of supporting tools and a restricted web page within BdP external site dedicated to non-financial institutions.