

**UNITED NATIONS
ECONOMIC COMMISSION FOR EUROPE**

CONFERENCE OF EUROPEAN STATISTICIANS

Seminar on Statistical Data Collection
(Geneva, Switzerland, 25-27 September 2013)

Topic (i): Centralising data collection

**CREATING A DATA COLLECTION DEPARTMENT: STATISTICS
PORTUGAL'S EXPERIENCE**

Working Paper

Prepared by Paulo Saraiva dos Santos and Almiro Moreira, Statistics Portugal

I. Introduction

1. Statistics Portugal (INE) started in 2004 an integrated and process driven approach aiming at re-engineering the production architecture, improving its efficiency and flexibility.
2. Consequently, a central data collection department was created, regional directorates were extinct, and domain production departments have been merged into three units: economics, social and national accounts. Methods and information system were merged into one department.
3. Survey's data collection is a core function of Statistics Portugal, consuming around 40% of its annual budget and 30% of its human resources. A Data Collection department assures mainly the operation of statistical production phases of collection, processing and analysis of collected microdata, covering all business and social surveys. Data collection staff is distributed all over the country (mainland and islands), especially in Lisbon, Oporto, Coimbra, Évora, and Faro, but under a centralised coordination. The Autonomous regions of Madeira and Azores have their own authorities for the production of regional specific statistics, while being the data collection centres for those areas for Statistics Portugal, under common technical requirements and infrastructure.
4. The benefits and lessons learned are substantial, especially on cost reductions, improvement of the quality. There are still many opportunities for further developments.
5. This paper offers a step-by-step historical view of this journey and lists the results achieved.

II. Statistics Portugal's background

6. Statistics Portugal is the central national authority for the production of official statistics. It aims at developing and supervising the national statistical system.
7. Created in 1935, Statistics Portugal has its head office in Lisbon with delegations in Porto, Coimbra, Évora and Faro.

8. Statistics Portugal is a public institution which has legal personality, administrative autonomy and technical independence in the exercise of its official statistical activity. It is a special public institution integrated within indirect State administration.
9. The Statistical Law confers on Statistics Portugal statistical authority and legal obligation to confidentiality. All surveys are mandatory and miss or delay response to surveys from statistical authorities is considered a serious administrative offence, submitted to financial punishment.
10. Its Statistical Operations are mainly conducted in accordance with European Council Regulations and other international duties. In 2013, 82% the activity of Statistics Portugal is ruled by European Union legislation and by the European Statistics Code of Practice

III. Reengineering the statistical production

11. From 1989 to 2003, Statistics Portugal had a number of regional offices outside Lisbon located, respectively, in the North, Centre, Alentejo and Algarve. Together they accounted for a significant percentage of the Institute's total staff, and the mission of each was threefold:
 - To support data collection, on the grounds that proximity to respondents allowed collection operations more effective;
 - To conduct at least one full subject matter programme with national scope, on the grounds that each region has its socio-economic specificity and it was needed a regional office to be sufficiently attuned to produce the corresponding statistics;
 - To contribute to the development of regional statistics, on the grounds that these are better handled out of a regional than a central statistical office.
12. This structure, based on alternative mini-statistical offices was considered inadequate, and could not be sustained in the long run.
13. It was a heavy and costly organization:
 - 788 workers: 37% in Regional Directorates.
 - 195 managers (25%):
 - 14 Departments, 5 Regional Directorates, 48 Units, 128 sections
14. There were high duplication of structure, work, procedures and tools were a consequence, promoting an institution not flexible enough for the future.
15. Like many other counterparts, Statistics Portugal produced statistics at that time through a non-integrated architecture. It was based on numerous parallel processes, domain by domain, according to a traditional stovepipe approach. This way of producing statistics was then considered complex, inefficient and not flexible
16. From 2003, a new board decided to take a proactive evaluation of the existing model, not guided by resources constraints. It was hired an external advisory company (international strategy consultants). Also, it was request a Peer Review in 2004, which was conducted by two reputed experts, Mr. Ivan Fellegi, former Chief Statistician of Canada from 1985 to 2008, and Mr. Jacob Ryten, former Assistant Chief Statistician of Canada from 1969 to 1997.
17. From the Peer Review recommendations are stressed the following topics.

- About responsibilities, Statistics Portugal should reshuffle responsibilities to the centre and to the other offices in a manner that combines two principles: (i) assigning coordination for all surveys with national scope to the center; (ii) while building on existing resources in the regions as well as in Lisbon to establish functionally centralized data collection functions in each;
- About human resources, to recognize that not all staff members are in a position to move from one urban center to another but that they all represent precious capacity. Accordingly, operational staff in either the regions or in Lisbon who cannot be fully employed in one of the new functionally centralized collection operations should be assigned to direct data collection from households or businesses in the regional offices concerned.

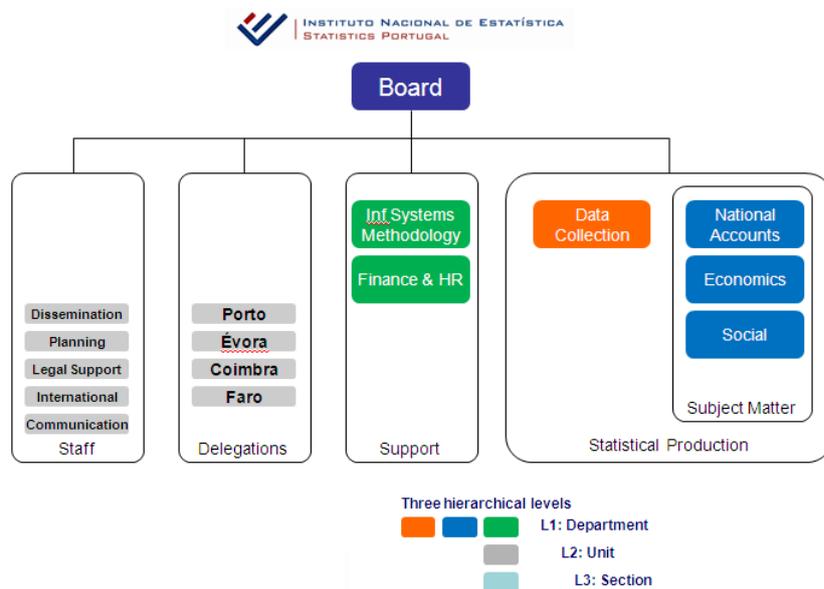
18. Based on the recommendations of the Peer Review, it was decided to create a new structure to Statistics Portugal. It was based on the following major changes:

- A central data collection department was created;
- Regional directorates were extinct;
- Domain departments were merged into three units: economics, social and national accounts;
- Methods and information system were merged into one department.

19. It was a remarkable challenge, considering the new distribution of resources, roles and responsibilities.

20. During the transition the statistical operations were not substantially affected, in spite of some resistances and other constraints.

21. Today, the simplified structure is shown as followed:



22. The following chart describes the present Production Architecture:



23. This Production Architecture has three subject matter departments been internal clients of a centralised Data Collection department, all of them supported by a single department assuring Methods and Information Systems competences.

24. The following chart describes the impact on the structure from 2003 to 2013:

	2003	2013	% Diference
L1: Departments	19	7	-63%
L2: Units	48	34	-29%
L3: Sections	128	13	-90%
Managers	195	61	-69%
Workers / Managers	4,0	10,9	173%
Workers	788	665	-16%
Lisbon	496	508	2%
Regions	292	157	-46%

25. A 16% reduction in the number of employees was mainly due to natural mobility and personal reasons. There was no pressure to reduce staff, neither structured programmes to promote workers to leave. The only exception was a programme to motivate regional staff to move to Lisbon, with a very low adherence. Of course, it happened a significant 46% reduction of the regional staff, because some workers did not accepted the new model and they moved to other institutions or companies.

26. One remarkable change was the 69% reduction of the managers, transforming a 4 workers per manager ratio to 10,9, which is considered a well balanced level.

IV. Centralised Data Collection

27. Survey's data collection is a core function of Statistics Portugal, consuming around 40% of its annual budget and 30% of the human resources.

28. A centralised Data Collection department assures mainly the operation of statistical production phases of collection, processing and analysis of collected microdata, covering all business and social surveys.

29. Data collection staff is distributed all over the country (mainland and islands), especially in Lisbon, Oporto, Coimbra, Évora, and Faro, but under a centralised coordination.

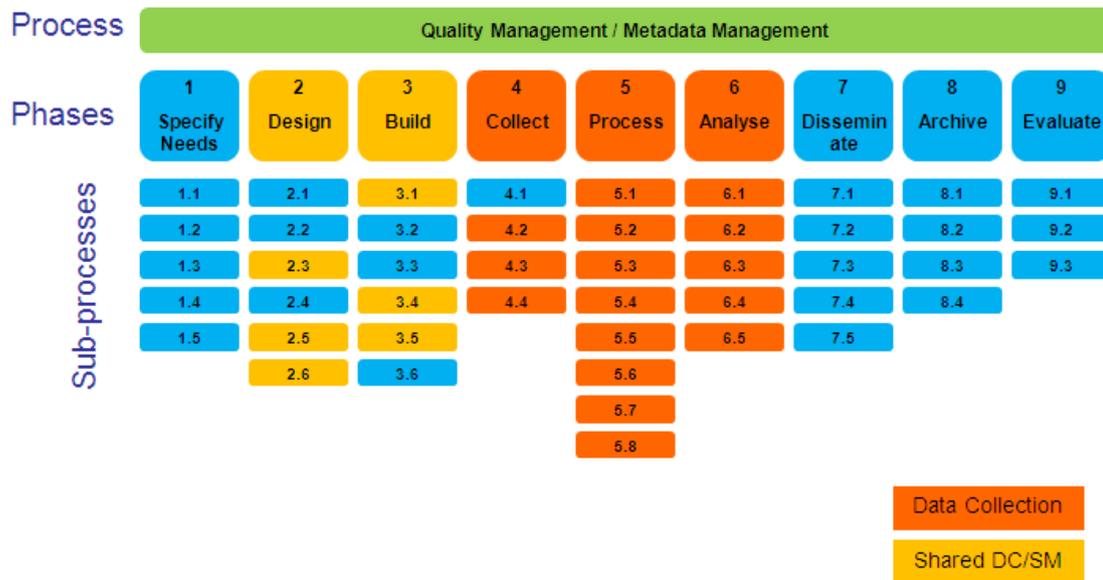
30. The Autonomous regions of Madeira and Azores have their own authorities for the production of regional specific statistics, while being the data collection centres for those areas for Statistics Portugal, under common technical requirements and infrastructure.

31. Statistics Portugal conducts 82 surveys in 2013, been 59 business surveys and 23 social surveys and price collection;

32. Data Collection department has 200 workers and 350 freelance interviewers. It is organized in three units and twelve sections.

- Self-completed surveys unit assures the business data collection and it is organized by project or statistical operation. Each of its six sections acts as the national management of the data collection activities of each survey;
 - Interview surveys unit assures interview and price collection. Each of its five geographical sections work with the same set of surveys, sharing same Share same methods, procedures and tools;
 - Data collection processes unit assures the national coordination for the data collection in surveys by interview. It also coordinates the activities of innovation on data collection methods, having one section responsible for the CATI national coordination.
33. The management of the data collection activities is in large extent decentralized inside de department, but with a strong centralised control. We foster a culture based on improving autonomy with clear responsibility;
 34. There is one overall budget for data collection distributed through each management level in the department. Also, central staff assures the control of the costs facing the estimated seasonality of the budget by survey.
 35. Another important too is the accounting the allocation of human resources to projects. The budget and the control are made by survey, project or activity, with a high level of detail, even in terms of a common set of tasks (e.g.. respondent help desk, data analysis, response follow-up, procedures development, project control, administrative support). There is a, individually daily allocation of the working time to each project code and task;
 36. This kind of time allocation tool was very useful in the early stages of the creation of the Data Collection Department, because this indicator (“data collection” resource consumption) was essential to diagnose the exact amount of resources used in this porcess, and support a creation of a centralised department. Also, the historical indicators were an excellent help that allowed the redistribution of the workers between the subject matter and data collection departments.
 37. Concerning the human resources management, a matrix approach is taken. It is very common workers participate in several surveys during the annual activity.
 38. The annual planning cycle begins with the Statistics Portugal Activity Plan. Then a Data Collection biannual activity plan is updated and all the resources needed are evaluated. Afterwards, the surveys and projects are distributed through the units and sections, having some workers that have their activities shared between teams.
 39. The relationship between Data Collection and Subject Matter departments was one major issue at the beginning of the re-engineering process.
 40. There were a negative perception of the Data Collection tasks, considered as “a low profile work ...”.
 41. Conversely, due to historical reasons, subject matter statisticians were very “data collection oriented”, and the expectations are always high. There was a deep perception of the subject matter colleagues that its collection counterparts should do better than them (when they were responsible for data collection).
 42. It was a very interesting phase, where issues related with organizational behaviour were more critical than the technical ones.
 43. The solution of this kind of conflict was twofold. (i) the definition of the boundaries between Subject Matter and Data Collection; (ii) 45. the adoption of Service Level Agreements (SLA) to manage expectations and to build trust.

44. The boundaries between Collection and Subject Matter were based on an internal procedures matrix, which has close similarities to the Generic Statistical Business Process Model (GSBPM) model, which is seen as a flexible tool to describe and define the set of business processes needed to produce official statistics.
45. Considering the GSBPM structure, the existent distribution of responsibilities is shown in the following figure. Note that there are few phases and sub-processes that are shared between Data Collection and Subject matter Departments.



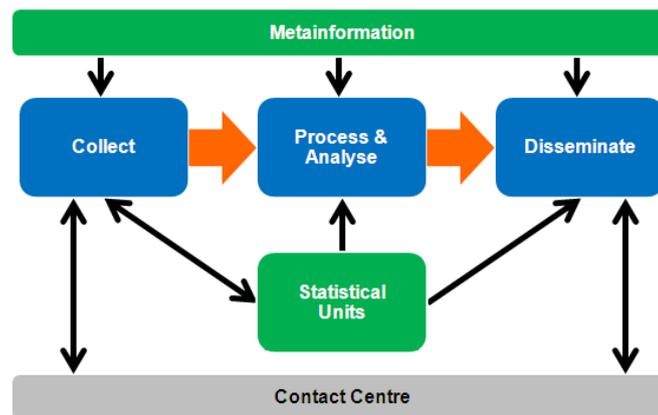
46. Concerning the adoption of SLA, we can mention that, at the beginning, we try not calling SLA to this approach. We found that the term has a negative perception and it does not help the negotiation; It was used a step-by-step approach, from a simplified version and increasing gradually the complexity.
47. Data collection SLA is survey specific. There are two main components: (i) a definition of the basic service level (included in the methodology of the survey); (ii) a variable data collection level, which is negotiated during the preparation of the survey planning phase. SLA just contains operational indicators, its measures, goals and intervals of acceptance and to overcome, and there are indicators continually measured during the operation phase.
48. Both tools are still considered critical instruments to the management, evaluation and development of the Centralised Data Collection model.

V. Data Collection infrastructure

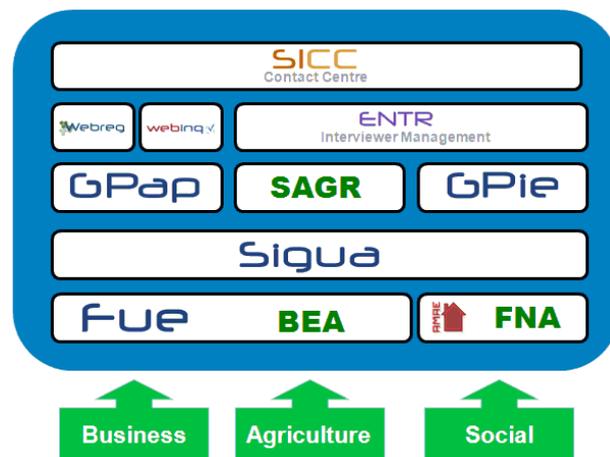
49. Integrating information systems is one major concern among companies and institutions, being widely accepted as a critical success factor of any organization.
50. The development and management of a common and integrated operational infrastructure is one of the most positive consequences of the centralised data collection approach.
51. A Survey Management System was developed and other Data Collection tools, such as the Datawarehouse and the Business intelligence solution (BIS) have been fully integrated into the operational infrastructure.
52. It began in July 2005, when the Internet service WebInq was launched, offering an easy and secure alternative to businesses to provide the survey data through electronic self-completed questionnaires.

After six years with the WebInq, more than 85% of the business questionnaires are collected electronically.

53. A new approach of statistical production was considered essential, focusing a broad integration, and process and tools standardization. Thus, a multidisciplinary working group was created in order to design an integrated architecture to support the production.
54. This effort resulted in an Integrated Survey Management System (SIGINQ), which covered firstly the business surveys, and later on the social surveys.
55. The SIGINQ aims at offering an integrated infrastructure to better support the statistical production and development in an efficient way, covering all the statistical operations (business and social). It unifies the main components into a comprehensive and interdependent system based on the architecture illustrated in the followed figure.



56. The system follows the basic production GSBPM phases, process, analyse and disseminate. Statistical units registers and metainformation support the flow of the processes. A contact centre system offers the infrastructure to telephone interviews (for social surveys), and the support to data providers.
57. The process management component is responsible for the management and control of all data collection processes, including information about respondents and paradata. These processes are full supported by the Metadata System. Process Management is subdivided into three other components: (i) GPap (self-completed surveys), for business surveys; (ii) GPie (interview surveys), for social surveys; and (iii) SAGR, for agriculture surveys. These three different components have similar features and functions, but adapted to each kind of statistical unit
58. The figure below shows the building blocks of a complete representation of SIGINQ. We can note that the system is specialised in three domains: (1) Business Surveys; (2) Social Surveys; and (3) Agriculture Surveys.



VI. Benefits of Centralised Data Collection

59. From our experience we can highlight following perceived benefits of the centralised data collection.
60. The Development and management of a common infrastructure, both intellectual and operational, which could only be duplicated geographically at great cost;
61. Creation of a flexible, dynamic and responsive production architecture tied to the common services provided by shared means of production (sampling frames, classifications and standards, questionnaire designs, methods and tools, etc.);
62. Creation of right means of coordination to make our design work in order to face future (but now present) budgetary cuts;
63. Adoption of a cost-effective approach that makes the most effective use of regional and central resources. It was possible to do more with the same.
64. Reduction of the data collection cycle, specially the time to deliver statistical results;
65. Assistance to develop a steady culture based on efficiency and innovation, considering the full in-house design and development approach;
66. Development of analytic competences in order to improve the quality of the information (more reviewing and validation tools);
67. Creation of an integrated Survey Management System as well as other Data Collection tools;
68. Reduction of respondent burden:
69. Avoiding duplication of variables and offering easy and multiple ways to provide data;
70. Reduction of production costs, e.g. estimated in 27.2% (business surveys; 2005 – 2012).

VII. Future of Data Collection

71. As long term goal we foster to improve our single collection platform for all surveys: business, social and prices.
72. During the last eight years we witnessed remarkable improvements in new data collection methods and modes. Business surveys have major benefits, including savings and quality improvement. Social

surveys have benefited the introduction of multimodal collection capabilities, but there are high expectations on the offer of web questionnaires to the families.

73. We identify the following topics and opportunities for our next steps in Data Collection:

- Increase the use of administrative sources;
- Extend Integrated Production Systems;
- Improve Automatic Data Transfer and the use of Scanner Data on price collection;
- Increase the multimodal collection capability (web based);
- Improve the use of paradata to support the quality processes;
- Create new processes to better understand respondent's behaviour in order to motivate their collaboration.