

**UNITED NATIONS
ECONOMIC COMMISSION FOR EUROPE**

CONFERENCE OF EUROPEAN STATISTICIANS

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

Topic (ii): Managing data collection functions in a changing environment

**DATA COLLECTION STRATEGY IN THE NEW INFORMATIONAL
ENVIRONMENT**

Working paper

Prepared by Nana Aslamazishvili, National Bank, Georgia

I. Introduction

1. We all are witnesses of the terrific speed with which the world economy and society as a whole are being changed. Correspondingly, an interest in statistics, which is designed to reflect these changes, their progress and outcomes, is growing at the exponential rates. Moreover, it is obvious, that traditional survey-based data collection becomes more and more obsolete nowadays. All this calls for strengthening a statistical capacity and directing it towards a production of more reliable and timely information. However, the finding a better balance between the users' needs and statistics is unthinkable without changing data production environment in a broad sense. In this context, Central Banks, like many NSOs, need to rethink own traditional data production framework and design a new statistical business process strategy.

2. The purpose of this paper is to discuss recent ideas and approaches how to new way of thinking and innovations, using new methods and technologies help to build a basis for re-engineering central bank statistics whole system and streamlining statistical business process. It gives an overview of main features and advantages of NBG's new innovative solution to the statistical business process organization, including data collection mode changing, collection process standardising and centralising, and automation of data production procedures.

**II. Data Collection Strategy as a Part of New Statistical Production
Conceptual Framework**

A. Existing Practice

3. Following the international statistical standards, NBG produces the majority of the country's most important official statistics (Monetary and Financial Statistics, Balance of Payments Statistics, IIP, External Debt Statistics, Interest Rates Statistics, Exchange rates Statistics). However, the existing traditional excel-based reporting and absence of the automated data processing mode leads to great

difficulties of data management and dissemination. In General, data can be produced in many ways and each mode has a different cost structure, advantages and disadvantages. It is worth emphasizing, however, that traditional reporting mode leaves much to be desired.

4. NBG's current data collection and production system might be characterized as:
- (a) Excel spread sheets based reporting,
 - (b) Decentralised between different units,
 - (c) Unsupported data validation and processing by the compliant IT technologies, and hence
 - (d) Unaccompanied with a centralised data warehouse.

Decentralised and multi-dimensioned datasets led to weaknesses of appropriate metadata.

5. There are significant differences between coverage of balance sheet data and statistical reporting. Usually balance sheet data structure is more "conservative" and doesn't reflect some of new standards and requirements, such as a residency or sectorisation criteria, etc. Data "gaps" are mostly eliminated through statistical reporting, that, in fact, leads to data overlapping. Hence, a scale of reporting burden, routine of data validation, processing and dissemination is quite significant for both data providers and producers.

B. Strategic Vision

6. On today's information-rich society, when statistical agencies face to speed increasing demand on statistical information, proper formulation of data production strategy has become of the most important objectives. Moreover, by oneself, the statistical business process is a complex phenomenon. However, the sensible utilisation of modern information and communication technologies can make processes streamlined and resource saving. This very approach was chosen by the NBG for transforming of existing business processes of the statistical production.

7. At the very beginning stage of our statistical production transforming process there were several complicated questions to be resolved:

- (a) Statistical needs: What would be more comprehensive set of data according to the international standards we want to have from financial institutions;
- (b) Collection strategy: What kind of mode would be more reasonable to change excel spread sheet based reporting over to web-based collection, and how to collect different data by the standardised way to ensure centralised collection;
- (c) Streamlining: How to optimize standard Statistical Business Process Model in order to reduce reporting burden as well as data validation and production costs.

III. SebStat: A New Vision on Statistical Business Process at CB

A. What is the best way to define what to collect?

8. There is a range of statistics NBG is responsible for. The Organic Law on National Bank of Georgia states, that NBG is responsible for statistics on financial and external sectors of the country. In order to reply challenges we faced transforming our statistical system, we took a rout to the elaboration of comprehensive and unified Statistical Information System for NBG's statistical production, so called SebStat (NBG's Statistics), able to collect, validate, produce, disseminate, and store financial and statistical data through the different way, we used to applying before.

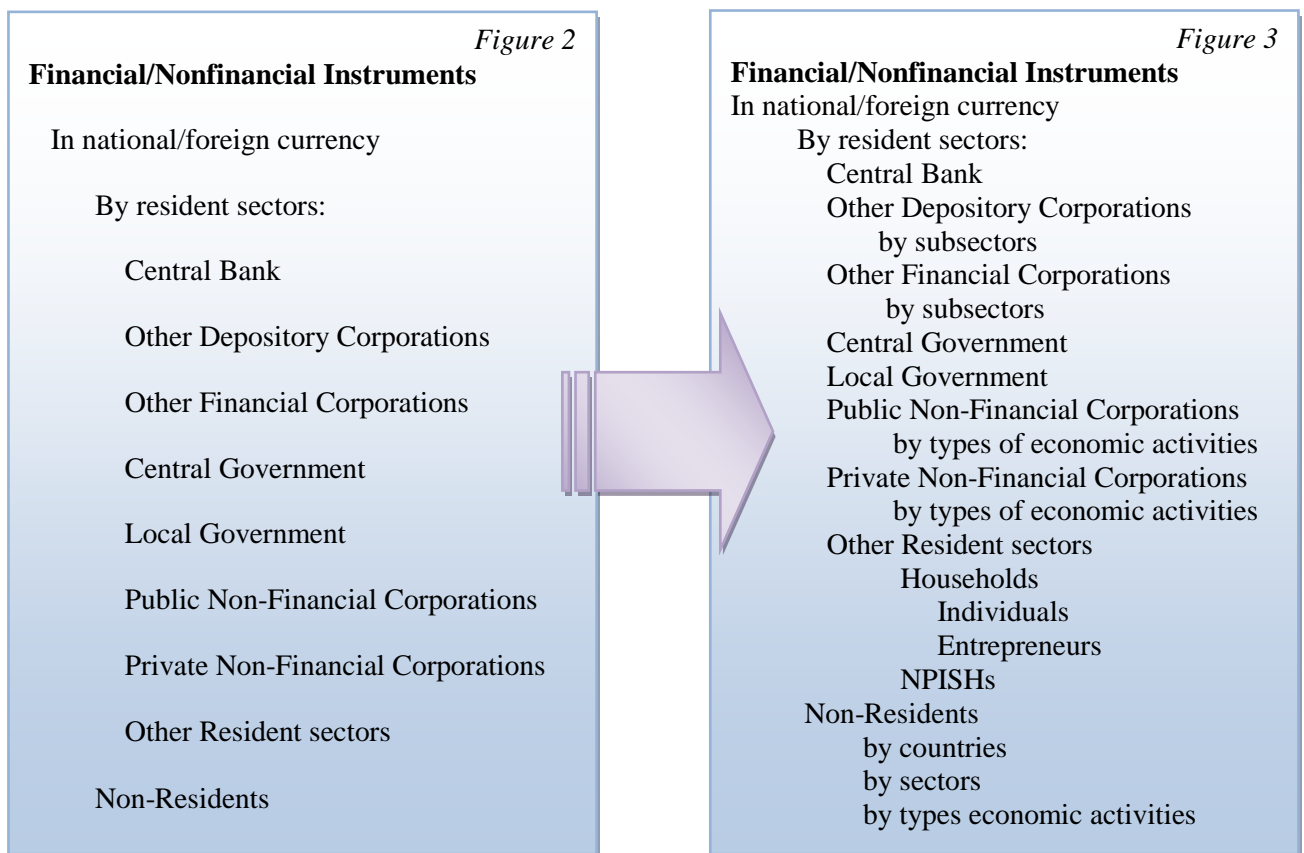
9. Central bank statistics need to cover a wide range of financial and non-financial instruments by the sectors and sub-sectors of the economy, types of economic activity, geographical areas within and outside of the country. The question that is most challenging for central bank statisticians is: How to define what to collect and what is the best way to do it?

10. To define properly what to collect we should know statistical needs.

Statistical needs, in terms of quality, availability, and analytical usefulness of financial and external sectors data, in both in-country and international context are clearly formulated in worldwide recognized manuals and guidelines, such as SNA, BoP, MFSM, FSI, and so on.

11. To be in conformity with international standards and integrally linked to other macroeconomic statistical systems, as well as to FSI and macro-prudential analysis, we decided in favour of balance sheet data structure for financial institutions, describing the unified framework for comprehensive analytical consideration.

12. *Figure 2* below shows structural breakdown requirements for financial data analytical purpose, used by the balance sheet approach, we are focus on. Moreover, we broadened this breakdown to deepen in-country analysis (*Figure 3*):



13. Hence, the structural features of the data to be collected are:

- (a) Primary breakdown by financial instruments;
- (b) Use of a balance sheet framework;
- (c) Instruments disaggregated by counterpart residency, sector, type of economic activities;
- (d) Breakdown by currency (national, foreign, and also by type of foreign currency, for BoP statistics and other needs).

Instruments disaggregation by appropriate features is also used (eg. loans by categories etc.), when it is needed.

14. Besides, important feature of the monetary and financial statistics today is necessity for collection not only of stocks, but of flows also, resulted from changes within a period of time:

- Transaction
- Revaluation

- Prices
- Exchange rates
- Other changes in volume of assets (OCVA).

SebStat as an information system is intended to reply all those demands.

15. What is the best way to collect data for balance sheets compilation?

The ideal way to collect data for balance sheet purposes can be easily found in a System of National Accounts fundamental idea regarding analysing flows and stocks, reflected under the question: “*Who does what, with whom, in exchange for what, by what means, for what purpose, with what changes in stocks?*” As stated in *2008 SNA* “Answering these questions for all economic flows and stocks and operators in a given economy would provide an enormous amount of information describing the complete network of economic interrelations“ (*2008 SNA*, Chapter 2. p.2.8).

16. In order to specify collection mode, analysis of the existing decentralised practices was needed. We had to review all data flows from banks delivered to the different units and investigate all possibilities for their integration into the centralised channel.

17. Hence, the initial goal of SebStat regarding data collection was three-fold:

- (a) Standardisation of data submission and production;
- (b) Centralisation of data collection, and;
- (c) Automation of all statistical business processes.

B. How to identify and describe inquiry for data providers by the standardised way?

18. Traditional excel spread sheets based reporting is understandable for human eyes and thoughts, but not for computer technologies. How to make requested data understandable for both – peoples and computers? How to standardise collection process for successful submission?

19. Initially, in order to make easy description of data we wanted to collect, splitting up the range of existing indicators by their contents was necessary. Consequently, we have formed statistical domains, reflecting different aspects of the domestic financial and external sectors activities, as well as most of the supervisory data. Each domain is associated with Data Families, such as:

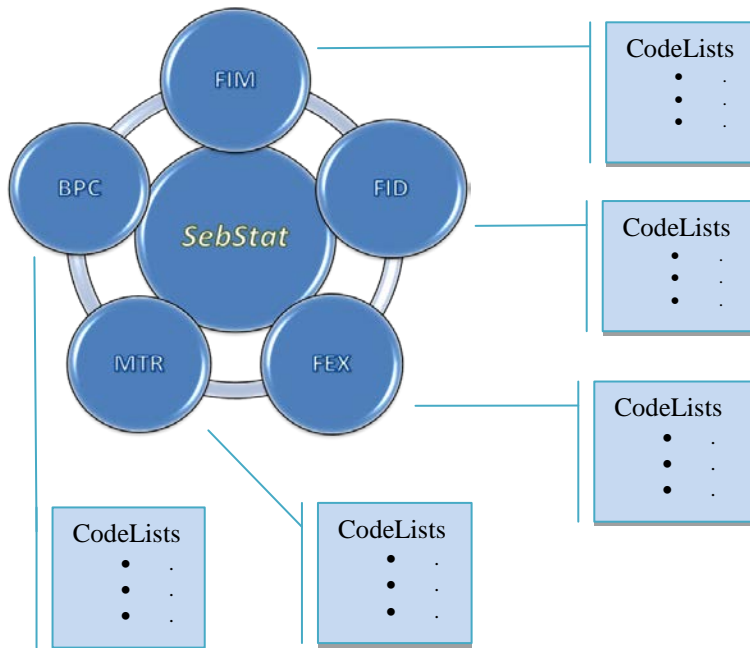
- (a) FIM – monthly financial and statistical data;
- (b) FID - daily balance sheets data;
- (c) MTR – money transfers statistics;
- (d) FEX – foreign exchange transactions;
- (e) BPC – bank payments cards statistics.

By the SebStat concept, each data family includes set of specific keys, identifying specific variables, banks are obliged to submit.

20. It is notable, that list of Data Families can be broadened in order to cover additional areas or needs, if it is impossible to do within existing domains.

21. In order to standardise data collection process, the method of coding of each requested data was utilized. Special Code Lists, created for each statistical domain (see *Scheme 1*) allow establish clear rules for data definition.

Scheme 1. SebStat Conceptual Framework



22. The *scheme 2* below illustrates how the data from existing excel spread sheet report can be identified in terms of SebStat, using appropriate Code Lists. This very approach is used for the data structure definition and request formulation for the reporting institutions.

Scheme 2. Identification of data structure from one of the traditional excel spread sheet reporting

Annual Interest Rate, %	Loans	Of which to:	
		Resident legal entities	Resident Households
Total	9,852,349	5,410,323	4,442,026
%	18.2%	18.1%	18.4%
16.0%	4,127,935	2,137,237	1,990,698
17.0%	432,433	281,656	150,778
17.5%	158,503		158,503
18.0%	2,153,075	1,329,062	824,012
19.0%	787,369	534,792	252,577

source frequency instrument currency residency sector Additional Info on loans collateralize Interest rate Data value

=

BBB.MM.F04000.GEL.RR.S14000.L010000.LC3.18.00.824012.00.20130531.

Completeness of the requested data depends on whether or not the appropriate Code Lists are well established.

23. The approach to data collection is similar for all Data Families. However, data structure is different for each one and is depended on the data features. As a consequence, NBG requires to banks just simple list of data separately for each Data Family, identified with specific range of keys (see *Scheme 3*), instead of excel-based questionnaires.

24. Data are loaded by the commercial banks automatically directly to NBG into the centralised data base. Data submission process is highly standardised and automated and this automation leads to significant reduction in reporting burden. Moreover, uploading process is accompanied with detail validation procedures. Data transmitted by banks via web-server are in XML format structured by the NBG.

25. In consequence, SebStat allows of optimize traditional statistical business process model considerably. From the respondents point of view that means no paper or excel spread sheets reporting, overlapping and data inconsistency. There is no necessity for data reconciliation among the different statistical domain, submitted to the different units of the NBG, as it was practiced before. From the NBG's side, standardisation and centralisation of data collection mode simplifies statistical needs formulation and demand designing procedures, and ipso facto leads to optimisation of the collection process substantially.

Scheme 3. Set of keys (fragment), identifying the required data

The screenshot shows the SebStat web application interface. The main content area displays a table with columns: ID, CL_Family, CL_Source, CL_Freq, CL_Ins, CL_Sta, CL_Stfl, CL_Mat, CL_Cur, CL_Res, CL_Sec, CL_Nac, CL_Reg, CL_Lop, CL_Loc, CL_Rng, CL_Ir, CL_Mtype. The table contains 15 rows of data. Above the table, there is a metadata section with fields: Organization (National Bank of Georgia), Creator_Name (naslamazishvili), and Creation_Day (2013-03-12T17:45:29). Two red callout boxes are present: one pointing to the table structure and another pointing to the first few columns of the table, labeled 'Keys, identifying the required data'.

ID	CL_Family	CL_Source	CL_Freq	CL_Ins	CL_Sta	CL_Stfl	CL_Mat	CL_Cur	CL_Res	CL_Sec	CL_Nac	CL_Reg	CL_Lop	CL_Loc	CL_Rng	CL_Ir	CL_Mtype
0	FIM	AAA	MM	F03104	LI	ST	XXXX	GEL	RR	S12100	JO	R00	XXXXXXXX	XXX	XX	XXXXXXXX	MV
1						FL		FRC	NN	S13099	L0	R13					VC
2																	OC
3	FIM	AAA	MM	F03104	LI	ST	XXXX	GEL	RR	S12100	JO	R00	XXXXXXXX	XXX	XX	XXXXXXXX	VC
4	FIM	AAA	MM	F03104	LI	ST	XXXX	GEL	RR	S12100	JO	R00	XXXXXXXX	XXX	XX	XXXXXXXX	OC
5	FIM	AAA	MM	F03104	LI	ST	XXXX	GEL	RR	S13099	L0	R00	XXXXXXXX	XXX	XX	XXXXXXXX	MV
6	FIM	AAA	MM	F03104	LI	ST	XXXX	GEL	RR	S13099	L0	R00	XXXXXXXX	XXX	XX	XXXXXXXX	VC
7	FIM	AAA	MM	F03104	LI	ST	XXXX	GEL	NN	S12100	JO	R13	XXXXXXXX	XXX	XX	XXXXXXXX	OC
8	FIM	AAA	MM	F03104	LI	ST	XXXX	GEL	NN	S12100	JO	R13	XXXXXXXX	XXX	XX	XXXXXXXX	MV
9	FIM	AAA	MM	F03104	LI	ST	XXXX	GEL	NN	S12100	JO	R13	XXXXXXXX	XXX	XX	XXXXXXXX	VC
10	FIM	AAA	MM	F03104	LI	ST	XXXX	GEL	NN	S13099	L0	R13	XXXXXXXX	XXX	XX	XXXXXXXX	OC
11	FIM	AAA	MM	F03104	LI	ST	XXXX	GEL	NN	S13099	L0	R13	XXXXXXXX	XXX	XX	XXXXXXXX	MV
12	FIM	AAA	MM	F03104	LI	ST	XXXX	GEL	NN	S13099	L0	R13	XXXXXXXX	XXX	XX	XXXXXXXX	VC
13	FIM	AAA	MM	F03104	LI	ST	XXXX	FRC	RR	S12100	JO	R00	XXXXXXXX	XXX	XX	XXXXXXXX	OC
14	FIM	AAA	MM	F03104	LI	ST	XXXX	FRC	RR	S12100	JO	R00	XXXXXXXX	XXX	XX	XXXXXXXX	MV
15	FIM	AAA	MM	F03104	LI	ST	XXXX	FRC	RR	S12100	JO	R00	XXXXXXXX	XXX	XX	XXXXXXXX	VC

26. Hence, SebStat is an innovative informational system which:

- provides a harmonized set of statistical concepts and terminology which is applicable across all statistical domain produced under the NBG's mandate;
- affords a number of opportunities for improving data management, provides a common platform for further development and implementation of standards for statistical data and metadata. It is a tool for knowledge and skills transfer to the financial corporations; allows financial Institutions

use it not only for data submission, but for own research and analysis also. At the same time, SebStat will be sufficiently flexible to cover swiftly innovations at the financial markets and reflect them properly in the monetary and financial statistics;

- helps both data providers and data producers to better understand their business from the methodological point of view as well as from technical one. Hence, it helps to increase the level of professionalism, on one hand, and improve the data quality, on the other;
- leads to reduce development and maintenance costs;
- makes easy to share experiences.

IV. Data Quality and Communication Strategy

A. Data Quality

27. Any effort directed to the improving and streamlining Statistical Business Process must be focused on data quality perfection. To address this challenge we prepared set of the methodological materials and technical tools within the SebStat Framework.

- (a) **SebStat Guide** is intended to explain to users the concept and methodology for producing internationally comparable data for submission to NBG. The concepts set out in this Guide are harmonized with those of the 2008 SNA, MFSM 2000, MFSCG 2008, BPM6, Financial Soundness Indicators, and Compilation Guide on Financial Soundness Indicators; **SebStat Guide** establishes a common terminology within statistical production process as well as between NBG and data providers;
- (b) **Code Lists** for each Data Family, with detail explanations;
- (c) **Data validation rules**;
- (d) **Bridge table** for identification links between of financial/nonfinancial instruments Code List and existing Chart of accounts;
- (e) **Case Studies** for clarification of different specific issues;
- (f) **Guideline** for data uploading;
- (g) Also, several NBG legislative decrees and other related documentation.

28. Special website, launched for SebStat needs, serves for interact directly with data providers and provide them related materials, good practice stories, news, recommendations and instructions concerning data quality improvement. SebStat forum helps to ensure prompt reply to the questions, data providers are interested in , and also to share experience among reporting institutions.

29. However, before launching of web-based tools for successful data submission, we carried out multitude of meetings and workshops with banks experts and IT specialists in order to ensure better understanding of SebStat conceptual and IT architecture, practical issues and communication strategy. It is worth to note, that SebStat implementation process plays important role in the creation of new vision to communication with financial institutions. Now it is more intensive, more effective, and more progressive then before.

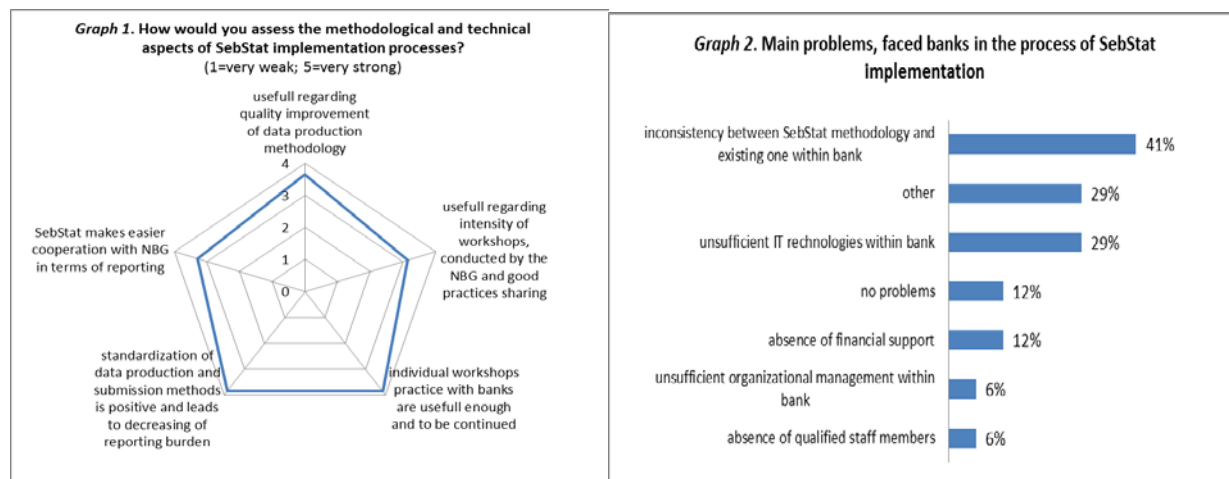
B. Choosing of New Communication Strategy: SebStat Implementation Assessment Survey

30. In order to make sure SebStat initiative was right project, we conducted a SebStat Implementation Assessment Survey among the project participant from the banks. The results are promising enough. On the other hand, they serve as recommendations for identifying priorities for future interactions.

31. The survey mainly was focused on cooperation quality assessment through sufficiency of the methodological and technical support, data quality strengthening and reporting burden declining expectation.

32. According to the responses of surveyed banks, SebStat facilitates the improvement of data production culture at the banks (66%), and would impact on data quality considerable (60%).

33. Most of the responses on the question: “How would you assess the methodological and technical aspects of SebStat implementation processes” (1=very weak; 5=very strong) was also positive (see *Graph 1*):



However, inconsistency between SebStat methodology and existing one within banks remains as a most pressing problem in the process of SebStat implementation (41%). The second pressing problem is inadequacy of existing IT technologies at the banks (29%). Probably because of this problem banks expectations regarding reporting burden are slightly negative (53%).

34. On the whole, banks consider that methodological and technical support from the NBG side is sufficient enough for SebStat implementation (4.29; scope 1-5) and would be still continued in the future. Moreover, 29% of surveyed banks consider that there is a room for future enhancing of SebStat framework.

V. Starting a Project with No Money: SebStat Curiosity

A. Budget Cuts and SebStat Perspectives

35. Budget cuts and statistical programs reduction are worldwide headache task now-a-days. NBG's statistics is no exception. However, starting a project with no money is a huge challenge.

36. At the very beginning of SebStat initiative, we faced two significant challenges:

- (a) Only verbal support from the NBG's high-level management, and
- (b) Inadequate number of staff in the Monetary Statistics Division, which was initiator of the statistical business process reforming at the NBG, and which is responsible for major statistical areas under the Central Bank mandate.

37. One successful businessman once said: “When you start a business, what must lead the way should be your own interest in or love for whatever it is you want to do. Loving what you do miraculously attracts all the necessary resources, people and opportunities”. Generally speaking, having no financial and human resources, this quotation became our philosophy regarding SebStat project development.

B. Prerequisites for change

38. Despite existing old-fashioned excel-based statistical reporting system, we mentioned above, during last decade we established stable statistical reporting system, with well-defined methodology in line with the international standards. Hence, talking with data providers about new initiative, capable to transform "stove-pipe" production into the modern automated business process was comfortable enough.

On the other hand, of course, the structural entities of the commercial banks, responsible for statistical reporting, are operating under different economic and institutional conditions (SebStat Implementation Assessment Survey illustrates it also), and the statistical reporting system management differs. That is why we faced a major challenge to find right way to motivate data providers think about re-engineering own business- and IT architecture in order to ensure SebStat implementation.

39. Motivation is essential for project success. We were unable to create motivating conditions for data providers. However, we managed to convince the banks of the need of SebStat for:

- (a) Significant reduction of reporting burden,
- (b) Automation of data validation and submission procedures,
- (c) Improving of data quality,
- (d) Increasing time for data analysis, and a lot more benefits besides.

40. One more element of project success is credibility and trust. In this regard, project team member's knowledge of international statistical standards, right and clear conceptual vision of the problem, standardised methodological and IT solution played crucial role. We had to change our communication strategy with banks dramatically in order to convince them, that SebStat project is accomplishable. In this regard:

- Since 2011 at present we carried out:
 - (a) 7 joint working meetings with existing in Georgia 20 banks representatives (with about 230 attended);
 - (b) 30 workshops with individual banks experts (with about 120 attended);

In addition to introduction of SebStat's conceptual and IT architecture, individual meetings with banks connotes also trainings and skills development, comprising a broad range of issues, such as statistical standards regarding residency and sectorisation criteria, CodeLists reconciliations, data structure definition and data validation issues, and a lot more besides.

- FAQs practice was implemented and banks are able to get detail explanations on all their questions;
- E-mail and telephone communications are very common for everyday interaction.

VI. Lessons Learned and Concluding Remarks

B. Lessons Learned

41. Despite a fact, that implementation of a new statistical informational system of NBG takes place in condition of extremely limited resources it is worth emphasizing, that SebStat is an ambitious project. However, it is successful thanks to number of well-considered points.

- (a) **Prerequisites:** first of all, it should be analysed all prerequisites (existing practice, professional experience and skills, data collection environment, etc.) for changes.
- (b) **Confidence:** to have a clear understanding of problem and ways how to solve it, before introduction of new idea, is essential in order to gain real trust and confident among stakeholders.

(c) **Priorities:** learn how to develop priorities, especially when resources are limited. It is essential to be sequent in actions.

(d) **Working Team:** Well organized working team plays important role for project success. Members of the team should be methodologists, analysts, national accountants, subject-matter statisticians (depending on Central Bank's mandate), bankers, financial accountants, IT specialists, programmers, depending on the nature of problems to be solved.

(e) **Long-Term Aims:** It is essential to focus own initiatives and ideas on a long-lifecycle statistical project; Statistics is too important to be taken lightly.

(f) **Cooperation Strategy:** Effective cooperation strategy with data providers and users is one of the important priorities of statistical production process organization.

(g) **High level management support** is essential for project success. It is crucial to understand, that "The only free cheese is in the mouse trap". It is essential to convince the management that it is the exact.

B. Concluding Remarks and Future Plans

42. Hence, it was our aim to create innovative informational system able to change dramatically "stove pipe" production into standardised procedures for each step of the data collection and production processes at the NBG. After the pilot round of collection, quality of data and uploading procedures are more than promising.

43. The adoption of the new centralised collection standards makes it easier to establish a flexible statistical data management system not only at central bank, but at the commercial banks also.

44. The main advantage of SebStat is that it makes much easier, less resource-intensive, and less burdensome on data providers.

45. SebStat as a system creates environment that facilitates enhancement of statistical data collection scope and quality, on one hand, and improvement of statistical capacity in terms of data presentation and dissemination, on the other.

As one of the next step of implementation, other financial institutions will be included in SebStat project.

46. In a broad sense, SebStat as a statistical informational system connotes formation of the Integrated Meta-Informational System. It is essential to analyzing and interpreting SebStat-related data and to making meaningful decisions.

Creation of the Integrated Meta-informational System will be one of the important achievements of the SebStat project.

47. It is worth emphasizing, that SebStat, by its concept and architecture, intended to be a system with a long lifecycle. On the other hand, data production now-a-days requires significantly higher skills and professionalism, then before. So, it is crucial to establish a good, continuing knowledge transfer mechanism in order to ensure using of system's capacity in a workmanlike manner.

Elaboration of comprehensive training materials for data providers, as well as for data producers and users is very important, and its realisation is one of important part of our plans in the near future.

48. Generally speaking we strongly believe that SebStat is the right direction to right positioning on the modern information market, and our strong expectation is that with SebStat we can produce timely, relevant and competitive statistics.