

Distr.
GENERAL

CES/AC.71/2005/30
22 March 2005

ENGLISH ONLY

**UNITED NATIONS STATISTICAL COMMISSION and
ECONOMIC COMMISSION FOR EUROPE
CONFERENCE OF EUROPEAN STATISTICIANS**

**EUROPEAN COMMISSION
STATISTICAL OFFICE OF THE
EUROPEAN COMMUNITIES (EUROSTAT)**

**ORGANISATION FOR ECONOMIC
COOPERATION AND DEVELOPMENT (OECD)
STATISTICS DIRECTORATE**

Joint ECE/Eurostat/OECD Meeting on the Management of Statistical Information Systems (MSIS)
(Bratislava, Slovakia, 18-20 April 2005)

Topic (ii): Development strategies for statistical information system

IT SITUATION AT THE STATISTICAL OFFICE OF SERBIA AND MONTENEGRO

Supporting paper

Submitted by the Statistical Office of Serbia and Montenegro¹

I. SHORT HISTORY

1. Relatively recent events in the territory of the former SFRYⁱ had, of course, a very deep impact on the whole situation in the statistical systems of all the involved parts. This paper offers a short overview of the effects of these events on the Statistical Offices, and focuses especially on IT.

A. IT in the Statistical system of SFRY and FRY

2. The Statistical system of SFRY was a centralized system, with the Federal Statistical Office responsible for data processing, analysis and data publishing at the SFRY level. The offices in the Republics were mainly in charge of data collecting, data processing and publishing activities at the level of the respective Republic.

3. The first computer in SFRY was installed in the Federal Statistical Office. From that day on, the whole Statistical system was based on the IBM mainframe technology. All Republic Offices were connected to the FSO mainframe, where the final dissemination of statistical data was made for the entire country. As the Republic Offices gradually installed their own mainframes, some dissemination was done at their offices and now, most of them are doing the whole dissemination by themselves after the SFRY broke apart.

4. The Statistical system of FRY was inherited from SFRY. It remained centralized, with FSO as the main headquarters for the two Republic Offices of Serbia and Montenegro. There were two IBM S390 mainframes (in FSO and RSOS) and an IBM Series 1 in Podgorica, which was replaced (when it stopped working) with just a control unit with several terminals to make a connection to FSO.

¹ Prepared by Branko Jirecek (jirecek@szs.sv.gov.yu).

B. Changes of the Statistical system in Serbia and Montenegro and their effects to the IT

5. By the Constitutional Charter of the State Union of Serbia and Montenegro, there is no real "federal government", so in that light almost all of the federal institutions were relocated to the republic level, with only a few remaining on the federal level (army, standardization, etc.). One of the "survivors" is the Statistical Office of Serbia and Montenegro, (SOSM) with responsibilities and obligations to collect, disseminate and publish statistical data for the State Union. The financing of all the State Union institutions was dramatically cut, which first affected the number of employees, which was reduced to merely 35 in the SOSM, from almost 200 in the former FSO. The budget of the new SOSM left absolutely no place for maintaining the IBM S390 mainframe. Due to all of this, there was hardly a way that the Statistical System would have stayed centralized, so it is slowly but inevitably becoming a decentralized system with more or less independent Statistical Offices of Serbia and Montenegro, and SOSM has been left with the only role of harmonizing and disseminating/publishing relevant data on the State Union level.

II. CURRENT IT SITUATION

A. Serbia

6. The main IT problem in the Statistical Office of the Republic of Serbia (SORS) is whether to remain on the mainframe platform or to move to a PC platform. The mainframe solution requires a relatively painful switch from RAPID to DB/2, while changing to PC looks even more painful.

Hardware

7. The SORS upgraded its S390 to a new IBM Z800 mainframe, completing the upgrading with new DASD units, cartridge units of greater capacity and other new peripherals. They have also improved their PC network. SORS has 15 branches covering all the territory of Serbia (without Kosovo). Branches have PC – networks or single PCs and are connected to SORS by dial-up modem connections. Exceptions are the Statistical Office of Vojvodina in Novi Sad with a UNISYS platform connected to RSOS mainframe with a leased telephone line and the Office for Statistics and Information technology of the City of Belgrade, with their own IBM mainframe, also connected via a leased line to RSOS.

Software

8. Data entry in SORS is mostly done in the branches and usually done with Blaise - written applications. The transferred data are entered, logically controlled and cleaned with GODAR, an application based on RAPID database developed and maintained entirely by the RSOS. GODAR is also used for some dissemination purposes. Unfortunately, the whole system is a bit outdated, and people in RSOS are very seriously thinking about changing it, but, even more unfortunately, the vast majority of surveys is done this way, so a lot of work and effort is needed.

9. Some applications (foreign trade, business register) have been transferred from the FSO mainframe when the responsibility for these surveys went to the Republic level. Those applications are based on IBM's DB/2 database.

10. There is a lot of activity in the RSOS on the PC platform, and they are doing a lot in Visual Basic and SQL-Server. For instance, the data entry, logical control, data cleaning and some publishing of the 2002 Census was done on the SQL-Server platform.

B. Montenegro

11. The Statistical Office of the Republic of Montenegro (MONSTAT) is in an easier position than the SORS, because they have no doubt about the platform, for it must be the PC. Nevertheless, their main problem is that they were, for years, completely dependent from the Federal Statistical Office (FSO) and SORS, where most of their surveys had been (and some of them still are) done.

Hardware

12. The situation in MONSTAT is that they have to use very unfriendly terminals to connect to FSO's and SORS's mainframes through a control unit via a leased telephone line. MONSTAT now has an Ethernet-based PC network, and will be very soon connected to FSO's and, even more important, SORS's mainframe from every PC in the network, which will be a very big improvement in every way.

Software

13. For a long time, connection to the FSO (and through this, to SORS) mainframe was the most important thing in the Statistical System of Montenegro. They would just enter their data in already made mainframe applications, "clean" them and publish the results. Some time ago, they started to produce their own applications, for the time being mostly in Visual Basic on SQL-Server. Statistics Sweden is helping them a lot in this effort.

III. CURRENT ACTIONS AND PLANS FOR THE FUTURE

A. IT situation in SOSM – staff, hardware and software

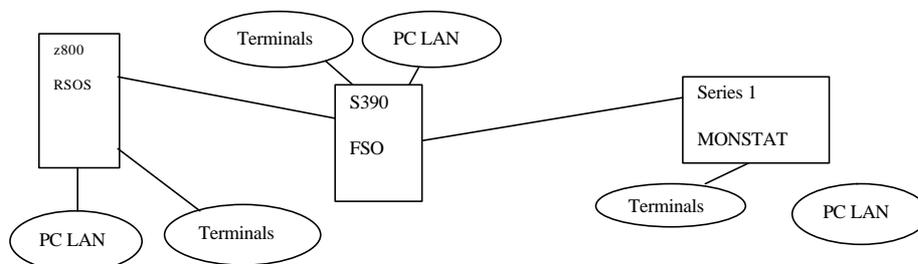
14. The above-mentioned reduction in financing for the State Union institutions created some very hard-to-overcome difficulties. The first is the radical reduction in the number of employees, from almost 200 to merely 35, the whole IT department being reduced to 4 people. Obviously, their work must be completely reorganized and done from the way it was when there were at least 60 people involved in the almost same amount of work.

15. The whole-year budget (whole, meaning not only for IT) of SOSM wouldn't cover even the software lease for the IBM mainframe, not to mention any maintenance or improvements. There were some negotiations with several institutions which we thought might be interested in "taking over" the mainframe so that we could both (SOSM and that other institutions) use it, but all of this was fruitless, so in the end, we will simply just have to shut it down. Our PC network is based on IBM's Token Ring technology, which we have decided not to change for the moment.

16. The software development was mostly based on IBM's PL/I and DB/2. In fact, all these applications are still running on our mainframe, because we simply don't have any alternative for the time being. The idea was, as could be supposed, to migrate all these applications to the SORS's mainframe, but the whole process is going very slowly and not at all to our satisfaction. We also have three PC's running Linux, installed for the Lotus Notes e-mailing system, connection to the Internet and the realization of the Intranet of the hardly existing Federation network.

B. Connecting the three offices

17. The connection between the two mainframes and a Series 1, later just a control unit in Podgorica was

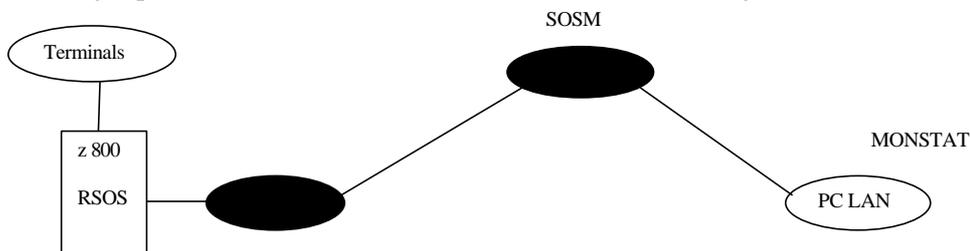


realized through IBM's control units and modems (communicating at an unbelievable speed of 9600 bps²) via the SNA protocol, as shown in Figure 1. Full communication was enabled, so it was possible to work on whatever machine from

² Believe it or not! Nevertheless, this speed was satisfying for most of the tasks.

any of the nodes in the network. This was also a way to solve the biggest problem - transferring enormous amounts of data.

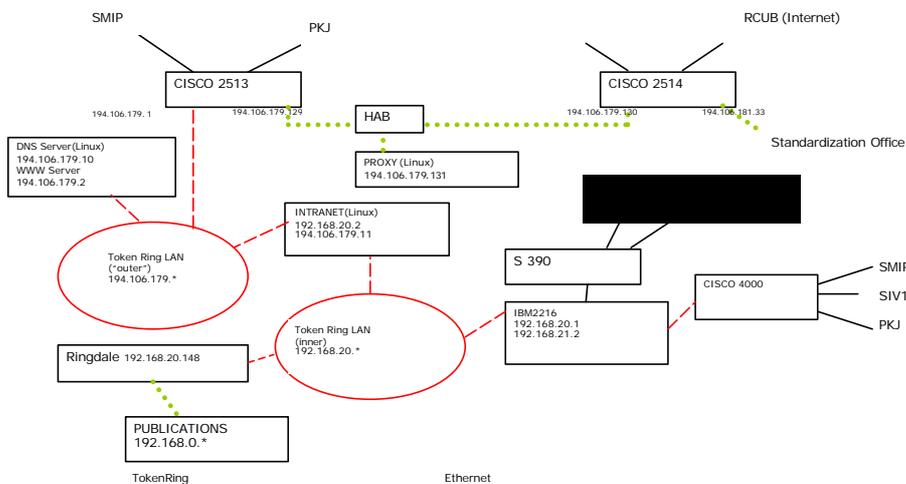
18. As for the inevitability of the “shutting down” of SOSM’s S390, all of its business is to be moved to the newly bought IBM z800 mainframe of the RSOS. The migration of data and applications is still on-going, and this alone could be an interesting topic to discuss about. This means that the functioning of the whole statistical system is going to change. SOSM is still supposed to collect, disseminate and publish data concerning the State Union, while the Republican Offices



are supposed to maintain data for the republic level. This is especially painful for MONSTAT, because a lot of their surveys were done at the FSO. Anyway, the main task is to preserve and save the functioning of the whole statistical system, so let’s try to deal with the problems instead of crying over the spilled milk.

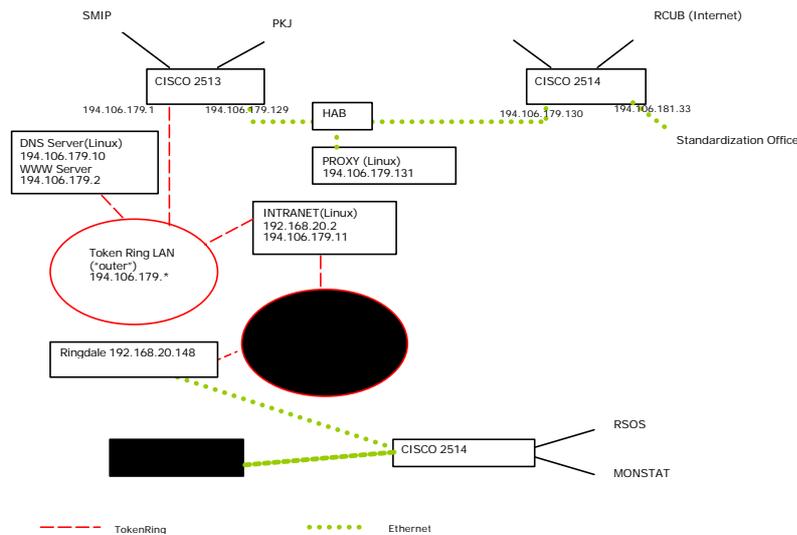
19. Obviously, the main problems are maintaining the functioning of the IT network between the three nodes and enabling data processing for the State Union to be possible without the mainframe.

20. As we saw it, the only possible way of connecting SOSM, SORS and MONSTAT is through the PC LAN’s that already exist in these three houses. The main task is to enable work with the SORS’s z800 from FSO and MONSTAT, because it is more than clear that this machine will have, in the first time at least, to do the work of the “deceased” SOSM’s mainframe as well as its own.



21. The connection should be something like it is shown in the figure above. The connections between the three nodes were realized through leased phone lines. The leased line in Beograd (between SOSM and SORS) proved to be capable of higher speed, so at this moment we managed to make an "alternative" link between SOSM and SORS, and, for the time being, it is functioning very well. The leased line to Podgorica, alas, is absolutely out of the question, so we had to lease a digital line to enable our communication. We expect that this connection will become operational in a matter of days.

22. As for the inner organization of SOSM, we now have the situation shown in Figure 3.



23. As we can see, the main net hardware in use is Token Ring, which is very unused nowadays worldwide. As the problem of changing net hardware to, for example, Ethernet, appears to be unsolvable, we are, for the time being, “condemned” to it. During the installation of printers in the publications department, we hardly found a Ringdale router that can communicate with Token Ring (our LAN) and Ethernet, on which these printers are running. As we already have it, we intend to use it as a “bridge” to the future connections to SORS and MONSTAT. We would simply use the existing Ringdale router, a Cisco 2514 router (bought in days of Federation network expansion), just buy 4 HDSL modems and, with some luck, the thing will be active. The new picture will look like in Figure 4. It should be noticed that the mainframe is not on this picture, and everything is happening without him. Even the former Federation network links from SMIP (Ministry of Foreign Affairs) and PKJ (Chamber of Commerce of Yugoslavia) to the S390 host are still “alive” through the Cisco 2513 connections (of course, there will be no S390, but their interest is data, not the machine), so that (almost) nothing is lost. The thing that we are mostly proud is that the whole operation, almost literary, doesn’t cost anything.

C. Organizing the Statistical Information System at SOSM

24. As said before, the work of statistical data processing was mostly based on the IBM mainframe technology, which means, in our case, ISPF environment, PL/I as the programming language and DB2 as the database. The whole LAN was designed, in the first place, to serve the mainframe, so little was done about the LAN itself. Not long ago, servers were running OS/2 operating systems. We did switch this some time ago with Windows 2000 servers with IIS (Internet Information Server), planning to give the SQL Server a “piece of the action”. This worked in some relatively satisfactory way, because the S390 is still active and the usual Windows bugs are tolerable.

Linux vs. Windows

25. The decision of the networking operating system between Windows and Linux still remains, with pros and cons for both sides. On one side, Windows is already running, so it means no work, just use it, not to mention that all of the workstations would definitely stay on Windows. On the other side, we already have three Red Hat Linux machines that are dealing with Internet, Intranet and Lotus Domino software and that have worked perfectly for all the time they are in use, Linux does look more like a real operating system that we are used to through the IBM mainframe(s) all these years, and is (still) free.

26. We further concentrated on the database question and we tested the (native) Linux PostgreSQL against SQL-Server. We imported all the data from SQL Server to PostgreSQL and (relatively) easy changed the existing Windows applications to “address” PostgreSQL instead of SQL Server – it worked quite well. The only “real” problem was the lack of unique identifier (auto-number, self-increment) field in PostgreSQL, which can

be relatively easy overridden with writing a little trigger function. PostgreSQL is more difficult to administrate, but is absolutely transparent from a Windows platform, so it could be easy administrated from, for example, Access. We didn't make no real "benchmarks", but through usual work, no one seemed faster (or slower).

27. In the end, we decided to stay on Windows. We seem to think that, in the end, if everything else is done properly, there will be very few problems to move from Windows to Linux at any time, so we decided to concentrate on the "real problem".

The (future) Statistical Information System

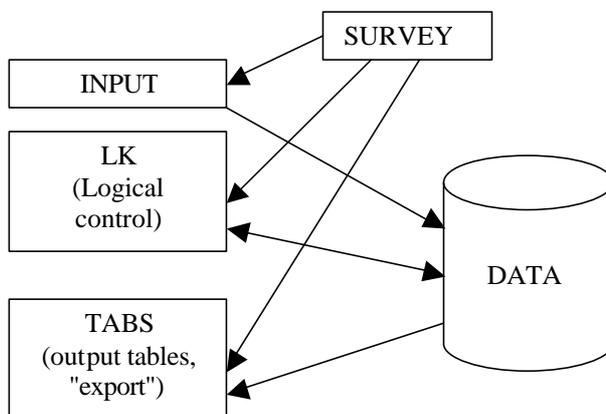
28. The main problem was to move the whole the mainframe-based environment for handling the data to a PC platform, keeping its functionality. The much bigger question was how to improve the overall organization and manipulation of the data, make them easier to administrate, disseminate, and publish.

29. As it has been said earlier, the DBMS system is, for now, the SQL-Server. We very quickly came to the decision that we will NOT use anything beyond the most usual and standard SQL. So, NO auto increments, triggers, functions etc. In this way we would be able to painlessly move to whatever relational-based SQL database in the future.

30. The choice of the programming language and environment was a much harder nut to crack. The thought of using some Linux tools had to be abandoned, because workstations and (even more) their users are "on Windows" and can hardly (if at all) be removed from it. One of the ideas was to try to make a compiler that would translate PL/I programs into some other language (Delphi seemed to be a good choice). This idea was abandoned when we made a little test – it took more time to "adapt" an output from our "stick-and-rope-made" compiler than to write the whole thing "from scratch". It is interesting that it works very good for some old (maybe even better to say: non-revised for a long time) projects, where the processing is more or less sequential, meaning without any real database at all – but these are just the things we really WANT to do from scratch, so there is no use of this again. As for our understaffed personnel, the C family was momentarily rejected, and so was Visual Basic because of our disastrous experience with it. The main candidates left were Delphi and Access, and we decided to do it a simpler way – using Access. In this way, the work of IT (in some way) doesn't stay just in IT, a little bit more experienced user (statistician) can also use all of the user-friendly Access approach.

Dealing with surveys instead of dealing with survey data

31. In fact, in every statistical survey you more or less have just three stages:
- Data Entry
 - Logical control
 - Tabulating (disseminating, publishing)



32. Very rarely, you have some calculation to do between the second and third stage, but we can omit this for a moment. Anyway, the whole idea is not much more complex even with that phase added. The idea is to organize the surveys themselves, not just to deal with their data.
33. This idea is not so new in SOSM (FSO). TABS, an FSO programming tool was, and still is in use for tabulating data. Modern report-tools (i.e. Crystal Report, even Access) are definitely much more powerful, but TABS was developed and made for a very unfriendly environment as IBM's ISPF is. Work on some kind of a similar tool for logical control was started several times, but never came to an end. It's not so bad – at least we have some knowledge about what certainly doesn't work.
34. For the INPUT phase, we started from an assumption that we will have to deal mostly with "flat" data (.txt files), as an expected output from the IBM mainframe. This will also be the most complicated situation, for everything else (i.e. Excel tables, .csv files etc.) could either be directly "transformed" or reduced to the .txt case.
35. The LK is the most difficult thing to achieve, because there are various levels of logical mistakes. The easiest is the "record" mistake, where the failure is in the record itself, and that one is easy to resolve with just interpreting a set of rules for that record. More difficult are errors that happen in the whole input table (duplicate keys as the simplest) or even between more than one table.
36. Assumption for the TABS phase is that every output is, in fact, a table. All we need for that table is a SQL query and a report. This, thanks to already mentioned report tools, turns out to be the easiest piece of the whole project.
37. The main thing that connects all of this is the code for every survey that we had in SOSM (FSO) for a very long time. The SURVEY set of tables consists of a range of information, starting from the name of the database and the database table name(s), security access to the methodology "manuscript" and to various metadata concerning the specific survey. This is supposed to be the master table(s) managing all the work with data in SOSM.
38. All the other stages are also just sets of tables, for instance, the TABS is just one main table consisting of the survey code, output table code, name of the resulting table, name of the SQL query and name of the report and a table consisting of some metadata for that particular output table (the survey metadata is in the SURVEY set of tables). The unique program "handler" will deal with all the surveys defined and handled this way. This way even more can be done – there is no real problem for crossing related data from different surveys, make web approach much more database-driven, and so on.
39. For the testing of this machinery, a survey from tourism was taken, to be more precise, this unhappy survey was a "test rabbit" for developing the bigger part of the machinery. I must say that we are very satisfied as for now.
40. It must be added that we do have luck that our mainframe is still functioning, and even when it "shuts down", we still have a chance to use the RSOS mainframe for all the time needed for developing this project. In that way, the production of statistical data will not stop, and we will just turn survey after survey into this way of handling.

41. We are aware that some of the surveys will not fit into. Some will require for some changes of all that was mentioned. We are hoping that none (or just maybe a few) will have to be standalone.

ⁱ Abbreviations:

SFRY	Socialistic Federal Republic of Yugoslavia
FRY	Federal Republic of Yugoslavia
FSO	Federal Statistical Office
SOSM	Statistical Office of Serbia and Montenegro
RSOS	Republican Statistical Office of Serbia
MONSTAT	Statistical Office of Montenegro
