Funded by the European Union

EUROPEAN UNION

COUNCIL OF EUROPE CONSEIL DE L'EUROPE

Implemented by the Council of Europe

Steering Committee of the International Monitoring Operation on the Population and Housing Censuses in Bosnia and Herzegovina

Seventeenth Assessment Report

21 – 23 April 2014

*GUIDO **PIERACCINI**, EXPERT*

*JEAN-MICHEL **DURR**, SENIOR EXPERT, SC TEAM LEADER*

Contents

List of acronyms

BiH	Bosnia and Herzegovina
BHAS	Agency for Statistics of Bosnia and Herzegovina
CoM	Council of Ministers of BiH
CMIS	Central Monitoring Information System
CSO	Civil Society Organisation
DB	District Brčko
EA	Enumeration area
FBiH	Federation of Bosnia and Herzegovina
FOS	Federal Office of Statistics (FBiH)
EI	Entity Instructor
EN	Enumerator
EUPHC 2	Technical assistance to Population and Housing Census Phase II
GIS	Geographic Information System
ICEI	Department for International Cooperation and External Information
IMO	International Monitoring Operation
IMO MG	International Monitoring Operation, Management Group
IMO SC	International Monitoring Operation, Steering Committee
IT	Information Technology
LFS	Labour force survey
MCC	Municipal Census Commission (Census Commission of the units of local self-government)
MI	Municipal Instructor
PES	Post Enumeration Survey
Q	Question
RS	Republika Srpska
RSIS	Republic Srpska Institute for Statistics
SA	Statistical Area
SI	State Instructor
TA	Technical Assistance
TAP	Technical Assistance Project
TOR	Terms of Reference

INTRODUCTION AND BACKGROUND

1. A Population and Housing Census shall take place in Bosnia and Herzegovina (BiH) in accordance with the Law on the Census of the Population, Households and Dwellings in Bosnia and Herzegovina in 2013 as adopted by the Parliamentary Assembly of Bosnia and Herzegovina on 3 February 2012. The Council of Ministers of Bosnia and Herzegovina has invited the European Commission to organise the international monitoring of the Census. Therefore, the European Commission, the Council of Europe and the Council of Ministers of Bosnia and Herzegovina signed on 18 April 2012 a Memorandum of Understanding to agree upon the following:

- The general objective of the International Monitoring Operation (IMO) of the Population and Housing Census in Bosnia and Herzegovina is to monitor the compliance of the whole Census exercise, from the preparation to the data dissemination, with:
- International standards on population and housing censuses as defined by UNECE and Eurostat, and as adopted by the Conference of European Statisticians as Recommendations for the 2010 Censuses of Population and Housing;
- Regulation (EC) No 763/2008 on population and housing censuses, and its implementing measures;
- The Fundamental Principles of Official Statistics, adopted by the UN Statistical Commission, as well as the European Statistics Code of Practice, promulgated by the European Commission;
- Standards on data protection and confidentiality, as provided for in the Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data of the Council of Europe, and the relevant European Union regulations in force.

2. The Operation will be carried out by a Committee of International Organisations, the Management Group (IMO MG) assisted by a Senior Census Expert, experts in population censuses, one or more experts in information technology and persons monitoring the census enumeration in the field.

3. The Senior Census Expert, the Census Experts and the IT Experts will assess the compliance of the Census in Bosnia and Herzegovina with the requirements listed in the Memorandum of Understanding, but they will not provide technical assistance.

4. Under the guidance of the Committee and the Senior Census Expert, the Census Experts shall:

- Examine the technical preparation of the Census including the pilot Census and post-enumeration survey, in particular the drafting of questionnaires and manuals;
- Monitor the collection, processing and dissemination of Census data, verifying the accordance with the requirements defined above;
- Control the fair and proper computation at all levels and whether the confidential nature of the individual data is fully guaranteed;
- Investigate any other matter with relevance for the Census.

5. The Management Group has designated Mr. Jean-Michel Durr as Senior Census Expert and Mr. Roberto Bianchini, Mr. Bent Noerby Bonde, Ms. Meryem Demirci, Ms. Kateri-

na Kostadinova-Daskalovska, and Mr. Guido Pieraccini as experts in population censuses. The Senior Census Expert and the Experts in population censuses form the IMO Steering Committee (IMO SC) of international experts.

6. The first mission of the IMO SC was conducted from 23 to 26 April 2012. The objective of this mission was to assess the preparations for the population and housing census in their main components, including the preparation of the pilot census. The report of the SC was sent by the Senior Census Manager on May 12 to the IMO MG.

7. The second mission was conducted from 26 to 29 of June 2012. The objective of this mission was to assess the progress made since the first mission in preparations for the population and housing census, including the preparation of the pilot census.

8. The third mission took place on 17 and 18 of September 2012 and was conducted by two members of the SC, namely, Mr. Jean-Michel Durr and Ms. Katerina Kostadinova-Daskalovska. The objective of the mission was to assess the implementation of recommendations made in previous reports and to assess the preparation of the pilot census, planned for October 15-29.

9. The fourth mission took place between 10 and 29 of October 2012 and was conducted successively by the members of the SC (5 days each, except the senior expert from 10 to 19 of October). The objective of the mission was to observe the conduct of the pilot census in the field. In addition, a team of four observers, composed of Lidija Naumovska, Gabor Rosza, Jean-Paul Sardon and Per Schöning was present during the whole period of the pilot census. Each of them covered several municipalities, in order to observe the pilot census field operations in all the 60 enumeration areas (EAs) sampled in the pilot.

10. The fifth mission took place from 19 to 21 of November 2012 (3 days) and was conducted by one member of the SC, Ms. Katerina Kostadinova-Daskalovska. The objective of the mission was to observe the conduct of the Post enumeration survey (PES) of the Pilot census in the field. The expert covered all 6 EAs (in 6 municipalities) sampled in PES sample in order to observe the PES field operations.

11. The sixth mission took place from 10 to 13 of December 2012 and was conducted by four members of the SC: Mr. Roberto Bianchini, Mr. Bent Noerby Bonde, Ms. Katerina Kostadinova-Daskalovska and Mr. Guido Pieraccini. The objective of the mission was to hear from the Statistical Institutions about the lessons learn from the Pilot Census and to assess whether the preparations are on level that could allow conducting a census in April 2013.

12. The seventh mission took place from 12 to 15 February 2013 and was conducted by one member of the SC, Mr. Roberto Bianchini. The objective of the mission was to assess the situation on the progress of preparation and implementation of the IMO recommendations in the new context of the postponement of the census, with a focus on the achievement of the deadlines regarding cartographic preparations. A second objective was to collect information and materials for a more detailed assessment on the progress of census preparatory activities to be made by all members of the SC in March 2013.

13. The eighth mission was conducted by the complete SC from 18 to 22 March 2013. The objective of the mission was to assess in detail progress made in the preparations with regard to the milestones given after the sixth mission in December 2012, when the SC recommended to postpone the census to October 2013.

14. The ninth mission was conducted by two members of the SC, Jean-Michel Durr and Katerina Kostadinova-Daskalovska from 22 to 25 April 2013. The objective of the mission was to assess the progress made in the preparations with regard to the milestones given in the sixth mission and updated in previous missions.

15. The tenth mission was conducted by the full SC from 28 to 31 May 2013. The objective of the mission was to assess whether the progress made in the preparations was sufficient to consider that the country was ready to conduct a census in October 2013.

16. The eleventh mission was conducted by three members of the SC, Mr. Jean-Michel Durr, Mr. Bent Noerby Bonde and Mr. Guido Pieraccini from 16 to 19 July 2013. The objective of the mission was to assess whether the preparations were still on track after the positive assessment of the 10th mission to conduct a census in October 2013. A special focus was put on the preparation of data processing.

17. The twelfth mission was conducted by four members of the SC, Mr. Roberto Bianchini, Mr. Bent Noerby Bonde, Ms. Katerina Kostadinova-Daskalovska and Mr. Guido Pieraccini from 3 to 6 September 2013. The objective of the mission was to assess the state of play regarding to forthcoming census field activities (including PES) and data processing activities following the field activities as well as the overall census preparations in accordance to the Census milestones and SC recommendations and advices given through the previous missions and SC reports.

18. The thirteenth mission was conducted by five members of the SC: Mr. Roberto Bianchini and Mr. Guido Pieraccini from 27 September to 4 October, Ms. Katerina Kostadinova-Daskalovska from 4 October to 11 October; and Mr. Jean-Michel Durr and Ms. Meryem Demirci from 11 October to 18 October. The objective of the mission was to monitor the enumeration of the census, including its final preparations and final operations. During this mission, the SC was assisted by a team of 27 international observers, deployed all over the country.

19. The fourteenth mission was conducted by one member of the SC: Ms. Katerina Kostadinova-Daskalovska from 1 to 3 November 2013. The objective of the mission was to monitor the conducting of the Post-enumeration survey following the Census. During this mission, the SC was assisted by a team of 5 international observers, deployed all over the country from 1 to 10 November.

20. The fifteenth mission was conducted by one member of the SC: Mr. Guido Pieraccini from 17 to 18 December 2013. The objective of the mission was to assess that the data processing system was functional and in accordance with the SC recommendations.

21. The sixteenth mission was conducted by five members of the SC from 11 to 14 February 2014. The objective of the mission was to assess the starting of the data processing and the implementation of the SC recommendations.

22. The seventeenth and present mission was conducted by one member of the SC: Mr. Guido Pieraccini from 21 to 23 April 2014. The objective of the mission was to control that the data processing was working smoothly and that the timing for completion of the phase was in accordance with the SC recommendations.

23. The expert wish to express his gratitude to the directors and staff of the Agency for Statistics of Bosnia and Herzegovina (BHAS), the Statistical Institutes of the Federation of

Bosnia and Herzegovina and Republika Srpska (RS) for their availability and collaboration, and to the staff of Council of Europe for the organisation of the mission.

EXECUTIVE SUMMARY

24. The selection of appropriate premises for storing the census materials has been finalized in time, before the end of the fieldwork. It is the opinion of the SC that the selected place fulfils the requirements, including security: a private guard controls who is entering in the place 24 hours a day; the access to the archive is limited to the archivists and the door of the archive opens with fingerprint recognition.

25. The space in the processing centre is well distributed and separate places are dedicated to different activities: the archive room to store the census materials, the server room for the central server of the whole data processing system, a large space for the preparation of the scanner materials, a separated space for the scanning operation and a vast room for the completion operators, coders and manual data entry.

26. During the present mission a random check of the quality of the census data was performed: some households were randomly extracted from the completed municipalities and the data recorded manually compared with the images. Over 16 households were checked and 63 individuals, and the only differences that were identified where on a few characters in the textual fields. These results confirm the good outcomes of the test performed by the three institutions.

27. At the moment 34% of the material was already scanned and the scanning process is supposed to finish in about 4 months. The whole data capture process (including character recognition and control) is at 12% of the individual forms. From this trend the estimation for finishing all the work is about 6.5 months. This means mid-November, which is a bit late to produce the first publication by the beginning of 2015, considering the phase of data editing that follows the data capture phase.

28. Two different options are possible to rearrange the process in order to complete data capture by beginning of October. The first one is to increase slightly the working time with minor changes, while the second, more effective, is to reduce the number of questionnaires that go into Quality Assurance review. This new organization will start in the next few weeks.

29. The solutions proposed by the Agency to ensure security and confidentiality of data are efficient and adequate and fully implemented.

30. In conclusion, the SC considers that the data processing phase is progressing well and in accordance with the SC recommendation. In the coming months the SC will continue to closely monitor the census data processing to assess its compliance with international standards and best practices.

DATA PROCESSING CENTRE

31. The procurement procedure to select the premises for the census data processing operations has been finalized in time, before the end of the fieldwork. It is the opinion of the SC that the selected place fulfils the conditions for the storage and the processing of the census

materials, including security: a private guard controls who is entering in the place 24 hours a day; the access to the archive room is limited to the archivists and the door of the archive opens with fingerprint recognition. The space it is well distributed and has separate places for the archive, the central servers, the scanning preparation, the scanning equipment and for the completion operators. At the time of the mission all the network requirements was fulfilled, including the backup to the server located in District Brčko.

32. In the last month, some problems with the IT infrastructure occurred in the processing centre and one scanner broke. This situation led to some days of inactivity and to one scanner stopped for about three weeks. This infrastructure or scanner problems may lead to some substantial delay of the processing phase.

33. A request to the EU delegation for additional funds to cover the maintenance of the IT infrastructure in the processing centre and the acquisition of a brand new scanner was submitted by the Agency.

Recommendations

34. The SC strongly recommends that measures to cover the maintenance of the IT infrastructure in the processing centre be taken and recommends in addition the acquisition of a new scanner. Indeed the risk that problems in this area may lead to substantial delay of the processing phase it is not negligible.

E-Flow Platform

35. The recommendation provided by the SC in its 15th report not to perform any manual modification on the paper questionnaires before scanning was fully implemented.

36. At the moment 34% of the material was already scanned (32,000 P1 per day) and the scanning process is supposed to finish in about 4 months. The whole capture process (including character recognition and control) is at 12% of the P1 (25,000 P1 per day). With this trend the estimation for accomplishing all the work is about 6.5 months. This means mid-November, which is a bit late to produce the first publication by the beginning of 2015, considering the phase of data editing that follows the data capture phase.

37. A full test of the E-flow platform was performed by the three Statistical Institutes during the month of February. The results of this process are well illustrated in the document "Results of testing OCR processing system on the census materials", submitted to the SC in mid-March.

38. The first step of the test was the random selection of 10 EAs from the sample used for the PES survey. For these EAs, a double entry of the data was made through the application for manual entry. After the second entry a field-by-field comparison was conducted. In this way, entry errors were discovered and corrected and the final data was established. Afterwards, a complete OCR processing of the observed EAs was conducted. The data obtained through the E-flow system and through the application for manual entry was then compared on a field-by-field basis.

39. The data comparison was performed considering the absolute number of differences and the number of differences having a value of the Levenshtein distance less than 2. The Levenshtein distance is a metric for measuring the difference between two character strings.

Informally, the Levenshtein distance between two words is the minimum number of single-character edits (i.e. insertions, deletions or substitutions) required to pass from one word to the other.

40. Comparing the data after OCR automatic recognition with those from manual entry, the following results were obtained:

	Total no of fields compared	Absolute difference (%)	Differences having LD<2 (%)
CHECKBOX	92841	99,51	99,60
NUMERIC	23081	91,22	94,91
STRING	12096	66,28	93,33

The automatic reading was the worst in the case of text fields. Only 66.3% of complete text fields were automatically read completely correct. The Levenshtein distance is therefore more appropriate for measuring the efficiency of text recognition, where we see that the correct reading is 93.3%. This means that, in average, in 6.7% of the total number of fields it was necessary to modify two or more characters in order to fully harmonize the text with the values on the paper questionnaires.

41. Comparing the data after the Completion and Quality Assurance (QA) phases with those from manual entry, the following results were obtained:

	Total No of Fields compared	Absolute difference (%)	Differences having LD<2 (%)
CHECKBOX	92841	99,76	99,76
NUMERIC	23081	95,95	97,94
STRING	12096	95,92	99,30

These results underlined that there was an improvement of data in the text fields up to 95.9% in absolute terms or up to 99.3% in terms of Levenshtein distance. Improvements occurred in the numerical values and in the check box fields showing that the Completion and QA phases worked properly.

42. Comparing the data after the Supervision phase with those from manual entry, the following results were obtained:

	Total No Of Fields compared	Absolute difference (%)	Differences having LD<2 (%)
CHECKBOX	92841	99,76	99,76
NUMERIC	23081	99,76	99,84
STRING	12096	96,50	99,37

These results show a small additional improvement of data accuracy whether we consider the absolute differences or the one based on the Levenshtein distance. The accuracy of checkbox and numeric fields is always above 99.0% while for the text fields it is above 96.0%.

43. During the SC mission a random check of the quality of the census data was performed: 16 households were randomly extracted from the completed municipalities and the data recorded manually compared with the images. The selection of the EA's was according to the following scheme:

	Total no of EAs processed	Proportion	Selected EAs
TOTAL	**1399**	**1,000**	**16**
TOTAL FED	**505**	**0,361**	**6**
BANOVIĆI	139	0,275	2
BIHAĆ	366	0,725	4
TOTAL RS	**894**	**0,639**	**10**
BANJA LUKA	862	0,964	9
BERKOVIĆI	32	0,036	1

44. In the 16 households and 63 individuals checked, the only differences identified where on few characters of the textual fields. No problem was detected in checkboxes and numbers, even in cases in which the enumerator made corrections on the questionnaires. More in details, the following table shows the results of the check:

	Total No Of Fields compared	Absolute difference (%)	Differences having LD<2 (%)
CHECKBOX	504	0,0	0,0
NUMERIC	567	0,0	0,0
STRING	441	5,9	0,7

45. This confirms the good results of the test performed by the three institutions. These results also confirm that the whole process (Tile, Completion, QA and Supervision) is well organized and going smoothly, which is also the impression that the SC had during the observation of the work in the processing centre.

46. The SC notes that the Agency currently does not own the source code of the platform and is not able to do any change to the system without the help of the external company that developed the system.

Recommendations

47. Two different options are available to rearrange the resources in a way that all the process will finish by the beginning of October. The first one is to increase a bit the working time with minor changes on the break time and on the working hours. The second and more effective option is to reduce the number of questionnaires that go in Quality Assurance from 100% to 70%.

48. It is the opinion of the SC that the second option will be more effective without sacrificing too much of the data quality (see point 41 and 42). Indeed at the moment 25,000 questionnaires are processed daily both in completion and QA. Reducing QA of 30% (to 17,500 questionnaires daily) will permit to increase Completion of the same percentage (to 32,500 questionnaires daily), permitting to align the amount of questionnaires daily processed to the amount of questionnaires daily scanned. Considering that there are about 3,300,000 remaining P1 questionnaires, it is estimated that this option will reduce the remaining time to about 5 months to accomplish the whole work.

49. To avoid that these changes produce a bottleneck in the Supervision stage, the number of Supervisors should be increased by 2 in each shift.

50. The three Institutions agreed that this new organization would start in the next few weeks, putting also in place some instrument to evaluate the impact that the reduction of QA will have on the quality of the data.

CODING

51. A global strategy is developed on how to perform the coding of the three main international classifications adopted for the census (ISCO, NACE and ISCED) and an ad-hoc application called g-Code is under development to approach this issue.
52. The g-Code application for coding the occupation is ready and tested with the pilot census data. The application is composed of two parts. A first part is running in background and - on the basis of a thesaurus derived from the last Labour Force Survey – the application attempts to automatically code the wording of the answer. A second application is used to manually code the wordings that were not automatically coded.
53. The codification of Municipality, Settlement, Country, Citizenship, Ethnicity, Religion and Mother tongue will be also performed with the g-code application: the same approach used for the codification of the Occupation is adopted.
54. The validity between Municipality and Settlement will be checked at this stage.

Recommendations

55. The SC recommends starting as soon as possible to test the whole coding process, also estimating the time and resources needed.

DATA EDITING AND DATA CORRECTION

56. The application that performs contingency controls is now developed. This application checks inter-records consistencies (relation between P1, P2 and P3, relations with the

reference person, etc.) and identifies the usual resident population. Nevertheless, a review and a finalization of these rules should be performed, also taking in account that it is part of the bigger process of data editing.

57. A discussion was initiated between the three institutions on the opportunity to use some statistical methods for data correction. The option suggested by the SC of using the Fellegi-Holt approach and the related hot-deck imputation methods is still under consideration by the institutions.

58. To correctly perform data editing and data correction, the institutions believe they need some technical assistance from international experts who could help them in conceptualizing the work.

59. The solution proposed to avoid any alteration of data through automated correction of logical errors is to share the set of rules defined at least with a group of professionals coming from the three Statistical Institutions. Scripts for automatic correction of logical errors must pass substantial testing before being approved for use. The algorithms shall be tested on real census material coming from the whole country. Every change of census data performed through any application must be tracked into the Census database and should be reversible.

Recommendations

60. The previous recommendation to define an appropriate global strategy to approach the editing and correction of the Census micro-data is re-affirmed.

61. The SC should be able to assess, and eventually re-run on the initial Census database the set of rules defined for automated correction of logical errors. Concerning the correction of logical errors, the SC still strongly support the use of sophisticated statistical methods for data correction, like the Fellegi-Holt approach and the related imputation methods.

CENSUS DATABASE AND PROTECTION OF PERSONAL INFORMATION

62. Protection of personal information is a central issue. Personal information should be processed according to strict rules of individual data protection and used only for the essential statistical purposed planned within the census data processing. During the processing of census data, names could be useful for two main purposes: identify duplicate records and perform the record linkage of the census individual data with the PES records.

63. Disclosure of personal data can partially happen in different phases of the data collection and data processing but the two main situations in which this can happen systematically are: i) accessing the questionnaire images; ii) accessing the digital data in the Census database.

64. The solution proposed by the Agency to avoid situation of type i) is that only the E-flow platform will access to the images collections via an application embedded into the platform and not accessible directly by the operators. At the end of the E-flow process of a specific EA both the tiff and jpeg images generated by the scanning are encrypted while the original files are deleted.

65. The solution proposed by the Agency to avoid situation of type ii) is that all the Census applications encrypt personal data when writing them to the Census database. This encryption is performed with a key known only to application, not stored in database. Such a solution should prevent access to personal data in database from System administrators and any other users.

66. The solution proposed by the Agency to avoid any alteration of data in the Census database is to prevent any changes to the database not made through an application, and if not prevented, then discover them. For this purpose the Census database logs all the changes to a location where the database administrator cannot alter them. Moreover, to access the Census database three different passwords, known only to the representatives of the three institutions, are required.

67. The solutions proposed by the Agency to ensure security and confidentiality of data are efficient and adequate and fully implemented.

68. It was objected to the recommendation of the SC to destroy personal information from the database that Article 22 of the census law provides that Common database includes all collected and processed data and shall be a property of all three statistical institutions. However, Article 46 of the same law provides that destroying the Census material generated from the Census-related activities of responsible bodies and organisations in Bosnia and Herzegovina shall be carried out by the Agency, in accordance with the Law on the Archive Files and Archive of Bosnia and Herzegovina and regulations issued on the basis of this Law.

Recommendations

69. The SC reiterates its recommendation to the Agency to analyse the legal framework related to archiving questionnaires and files and to present to the SC a proposal of treatment of personal data including images, paper questionnaires and database.

70. The SC reiterates its recommendation that at the end of the data processing the images be destroyed and the personal data be removed from the Census database. Indeed, since the paper questionnaires will be archived, there is no need to maintain any digital version of them for historical reasons. Would it be needed, the SC considers it would be worth amending the census law to ensure the strict protection of individual data and protect the Statistical Institutions from any suspicion of misuse of data collected.

ANNEX. AGENDA OF THE MISSION

AGENDA

International Monitoring Operation Steering Committee

Date: 21 - 23 April 2014

Monday, April 21, 2014, Agency for Statistics of BiH (BHAS), Sarajevo		
09:00 – 13:00	**Meeting with the members of the IT working group from the Agency for Statistics**	**BHAS, Sarajevo**
13:00 – 14:00	**Lunch break**	
14:00 – 15:00	**Meeting with the members of the IT working group from the Agency for Statistics**	
15:00 – 16:00	**Meeting with Dijana Sikima from the EU delegation**	

Tuesday, April 22, 2014, Central Location for Data Processing		
9:00 – 10:00	**Meeting with the members of the IT working group from the three Statistical Institutions**	
10:00 – 10:30	**Meeting with the Coordination working group from the three Statistical Institutions**	
10:30-13:00	**Check the quality of the data processed**	
13:00 – 14:00	**Lunch**	
13:00 – 16:00	**Check the quality of the data processed**	

Wednesday, April 23, 2014, Central Location for Data Processing		
09:00–13:00	**Check the quality of the data processed**	
13:00–14:00	**Lunch break**	
14:00-15:00	**Meeting with the members of the IT working group from the three Statistical Institutions**	
15:00-16:00	**Meeting with the Coordination working group from the three Statistical Institutions**	