

TWINNING CONTRACT

Support to the State and Entity Statistical Institutions, phase V Bosnia-Herzegovina



MISSION REPORT

on

Promotion of Dissemination

Component no 3.3

Activity no 3.3.7

Mission carried out by
Jesper Ellemose Jensen

&

Annegrete Wulff
Statistics Denmark

16-7 – 19-7- 2013

Version: Final

		 STATISTICS DENMARK
--	--	---

IPA 2008

*Jesper Ellemose Jensen
Statistics Denmark, Sejrøgade 11
DK-2100 Copenhagen Ø
Denmark
Tel: +45 3917 3244
Email: jej@dst.dk*

*Annegrete Wulff
Statistics Denmark, Sejrøgade 11
DK-2100 Copenhagen Ø
Denmark
Tel: +45 3917 3155
Email: awu@dst.dk*

Table of contents

Table of contents	2
List of Abbreviations	2
1. General comments.....	3
2. Assessment and results.....	3
3. Conclusions and recommendations	7
Annex 1. Terms of Reference	9
Annex 2. Persons met.....	13

List of Abbreviations

BHAS	Agency for Statistics of Bosnia and Herzegovina
FIS	Institute for Statistics of Federation of Bosnia and Herzegovina
RSIS	Institute for Statistics of Republika Srpska
ToR	Terms of Reference
RTA	Resident Twinning Adviser

1. General comments

This mission report was prepared within the Twinning Project „Support to the State and Entity Statistical Institutions, phase V”. It was the sixth mission to be devoted to dissemination within Component 3.3 of the project.

The consultants would like to express their thanks to all officials and individuals met for the kind support and valuable information which they received during the stay in Bosnia-Herzegovina, and which highly facilitated the work of the consultant.

The views and observations stated in this report are those of the consultants and do not necessarily correspond to the views of EU, BHAS, FIS, RSIS or Statistics Denmark.

The concrete activities of the mission were:

- Follow up and expansion of the pilot installations made in December and January 2012
- Continued training in PX-Web (on the job training)
 - Test of the translated user interface
 - Web deployment process
- Continued training in PX-Edit for production of PX-files
 - Work with multilingual tables (4 or more languages),
 - Import of data from other files (File>Read),
 - Split and join tables
- Continued discussions on how population and housing census data can best be presented

As the main focus was the potential for electronic dissemination of census data using PX-Web the key expert on dissemination from the EUPHC2 project Mr. Bozo Grbic - participated in the workshop. As such the mission was a joint effort by the Twinning project and the EU project.

The mission took place in 3½ days. Due to the short time of the mission the experts find that the discussions were constructive. Due to other commitments Annegrete Wulff participated for 2½ day.

Before the mission it was the hope of the consultants that data from the census pilot could be used for trial preparation of PX-files. Do to various confidentiality issues this turned out not to be possible. The most important conclusion / recommendation from this activity is that the BC must start working on their own work using data from the population census pilot. This will clarify outstanding problems related to how the work is best organized among Dissemination, IT and Subject matter staff and help improve confidence in the software.

The twinning project still have the possibility of assigning more activities in this area before it closes down at the end of December 201, however additional needs can only be identified as the BC starts the implementation process.

2. Assessment and results

Follow up and expansion of the pilot installations made in December and January 2012

Since the last activity in December and January the user interface of PX-Web has now been completely translated from English to the relevant local languages. This part of the work can therefore be considered complete. The BC side provided valuable feedback on the process that we will now pass on to Statistics Sweden that is responsible for the development of the PX-Web software. The comments included ease of translation and time out problems.

The multiple language support of PX-Web is giving some problems with relevant languages in Bosnia-Herzegovina. These issues are technically speaking minor bugs but they will need to be fixed for PX-Web can be put into production. Statistics Denmark will ensure that these bugs are fixed as soon as possible. Although they are real issues – they are also fixable and will not hinder the dissemination of preliminary results of the census using PX-Web.

Test of the translated user interface

The translated user interface was not tested as such during the mission, but the BC side appears confident that the translation is valid and understandable for the users. Statistics Denmark will ensure that the translated files are made available to other potential users of PX-Web.

Continued training in PX-Web

The work done by the BC before this activity clearly shows that the PX-Web software is full and clearly understood by relevant staff. Changes to the look and feel of the software was briefly discussed but the relevant IT resources demonstrated in December and January that adapting this part of the software will not be an issue.

Multi-lingual files

For obvious reasons the dissemination will have to cover 4 languages. As previously stated a bug in PX-Web needs to be fixed before users can shift effortlessly between the 4 languages. An error report has been raised with Statistics Sweden who is responsible for the software. Statistics Denmark will try to ensure that a fix is available in August.

Web deployment process

As the PX-Web software's use and adaptation is clear to the BC. The web deployment process is also understood. The impossibility for connecting PX-Edit / PX-Web directly to a database containing individual records was discussed and explained.

Confidentiality issues

It was discussed if PX-web could handle confidentiality / discretion issues. Although a very limited functionality is found in the PX system it is not recommendable to use it. It is our experience / policy that confidentiality issues must be handled in advanced and by the responsible subject matter expert. It can and should not be done by IT and / or Dissemination staff.

PX-Edit has a facility designed to help "anonymize" data. However the functionality is limited to simple replacement of values inside a specified range with a fill character, i.e replacing all values in a specific range with a blank or "...". This however will only insure discretion if it the specific values can't be deducted from the overall totals.

In general the problem with confidentiality issues is that the data that is left out do to discretionary issues should not be available from the totals, thereby imposing an overall limit to the level of detail that can be reproduced in the database.

A general rule at statistics Denmark is that at least 5 observations are need in a cell. However when disseminating data at the settlement level this is most likely not enough to hide / mask small ethnic groups as these might be appear from the totals. In the theoretical example below it is clear that the entire distribution must be considered before publishing.

Population by ethnicity and settlement

Ethnicity	Number of Persons
Ethnic -1	100
Ethnic -2	96
Ethnic -3	4
Total	200

Illustration of tab to “anonymize” data

Variable: Federacija BiH / Kanton

Keywords Meta part View Select Sort Edit Add Anonymize

FEDERACIJA BiH
UNSKO - SANSKI KANTON
KANTON POSAVSKI
TUZLANSKI KANTON
ZENI_KO - DOBOJSKI KANTON
BOSANSKO - PODRINJSKI KANTON
SREDNJOBOSANSKI KANTON
HERCEGOVA_KO - NERETVANSKI KANTON
ZAPADNOHERCEGOVA_KI KANTON
KANTON SARAJEVO
KANTON 10

☐ Codes
☒ Other values
☐ Control values
☐ Zeros
☐ Dot codes
☐ Many controls
☒ separately
☐ sum
☐ product

Fill item: .. Mask range: 0 40 Anonymize

Table mask
Table: Mask

Continued training in PX-Edit

PX-Edit is required to produce the content (data files .px) that will be displayed to the users. Again the software is easy to understand and is clearly understood by the BC. The training challenge is clearly focused on the need to make the conversion from tables designed for printed publications to tables designed for electronic dissemination.

Think electronic – not paper

The layout of the proposed printed tables- to some extent- mix up variables and variable values. It is not clear to the consultants if this will work or not on paper. It is however an imperative that the tables designed for electronic dissemination is based on “cubes” where all variables can be crossed. It appears

that it is intended to publish paper publications broken down to the settlement level. It is our opinion that the cost and user friendliness of such detailed paper publications does not compare well with making data available electronically.

The PX database is meant for multidimensional tables where all variables are crossed. This gives a much richer level of information and also more flexibility for the end users to retrieve and build the needed information.

It is advised that the database does not depend on the publication tables but instead includes these crossed tables, of course with respect to confidentiality. To manage the confidentiality it is recommended to produce more tables where some does not include settlements, others excluding one of the variables, presenting the data at a higher level only

Focus on identifying variables

The main problem in producing / defining the files will not be the software. Instead it will be identifying the variables found in the required tables. The IT sectors may provide some help in this as it appears that some of this information is already found in suitable database formats from the pilot census.

Geographic conventions

The final dissemination will include more than 7000 settlements throughout Bosnia-Herzegovina. Various problems and methods of presenting this in a user friendly way was discussed during the training. The MS experts will explore one or two options further although these are more long term solutions related to the PX-Web software and are not likely to influence the initial dissemination. The existence of duplicates in the geographical naming can partially be overcome by disseminating data in a table for each Canton or perhaps even at the municipality level to avoid duplication of settlement areas. It is our opinion that showing the name in combination with a geographic code should be considered carefully as the coding may not be understood by the users.'

Illustration showing duplicity based on 1991 geographic list.

The screenshot shows a web interface titled "Search values for Settlement". On the left, there is a search section with radio buttons for "Code", "Value", and "Show all values". The "Value" option is selected. Below this is a text input field containing "Pale" and a "Search" button. Further down are two more radio buttons: "Search beginning of text" (selected) and "Search anywhere in text". A grey box contains instructions: "To select all values within the search result, click the top value, press the Shift key and then click on the last value". On the right, the "Search results (number of hits: 6)" are displayed in a table with two columns: "Code" and "Value".

Code	Value
112453	PALEZNICA DONJA
112461	PALEZNICA GORNJA
144606	PALE
148342	PALE
148369	PALEZ
159514	PALEZ

Creation of files

Various methods for building the files was discussed and demonstrated. It would have been more instructive to use population census pilot data but some conclusions are still clear.

The files can be created from scratch using PX-Edit and then adding the data later. Also excel can be used to create outlines / designs that can be used to create files in PX-Edit. Excel templates / examples with 2, 3 and 4 variables were prepared to illustrate how data is best structured for easy import to PX-Edit.

However the easiest way is to create data and the basic metadata (heading, variables and variable values) directly from the database containing the “cleaned” census data. This can be done in corporation between IT and Dissemination. A demonstration on a laptop on Danish dummy data using a MS-SQL database holding 11 million records shows that CSV / Excel files for initial creation (including aggregation) of a PX file can be done in less than 45 seconds. This clearly shows that the files can be created through “try and error” by dissemination and it staff working together. Only when sitting down together creating files will the outstanding issues become clear. This should therefore happen as soon as possible.

Creation of files – Production phase

For a continuous update of many files it is worthwhile setting up a program retrieving data direct for the micro database and aggregate to the files going to PX-Edit. In PX-Edit the needed metadata information is added (mandatory key words).

ELIMINATION is an optional keyword that in PX-Edit is marked at the Variable level. If marked with Yes it means that the user does not need to select anything for this variable, the sum of all the values will automatically be calculated. If a “total” figure is found in the data this is used as the ELIMINATION value.

Limit the number of variables

The expert created a set of templates giving the format input to px-edit if the file consists of 2, 3 or 4 variables. The technical limit of the number of variables in PX-web is 13. Nevertheless the experts recommended not to exceed six variables.

3. Conclusions and recommendations

As this was primarily a training mission it is difficult to give specific recommendations. However the following recommendations can be provide:

Dissemination and IT should start production of files using pilot data as soon as possible- this is the only way to identify outstanding issues

The printed books being prepared to publish census results can provide valuable information on which tables to produce – but should not be reproduced 100% instead the advantages of electronic publishing should be used through the publishing of multidimensional cubes.

PX-Web should be used to disseminate not only census data but also data from the normal statistical production in Bosnia-Herzegovina. Statistics with a high frequency of publishing should be prioritized i.e monthly statistics. Especially foreign trade statistics are interesting as detailed level data is expensive to produce on print and much easier to use when available in electronic format.

Representatives from Bosnia and Herzegovina should participate in the annual PC-Axis meetings – The next reference group will be in Albania 2 and 3 October 2013.

4. Next Mission

The next mission in component 3.3 is currently not scheduled. The twinning project is expected to be extended into December 2013. However it has been agreed that no activities will be conducted during the census period.

According to the twinning project work plan the outstanding activity 3.3.8 is a general activity to complete various outstanding issues regarding dissemination. It is our recommendation that the mission is scheduled in November /December. According to current planning it will be carried out by two experts either Ellemose Jensen and / or Knudsen and / or Wulff.

If it turns out to be required or otherwise fruitful it is suggested that the activity is not finally scheduled before it is known if the implementation of PX-Web in relation to the census requires additional work.

If it is turns out that there is an urgent outstanding issue a short (1½) mission can be scheduled by a single expert with short notice.

Annex 1. Terms of Reference

Dissemination
16th July 2013 to 19th July 2013
Terms of Reference - Activity 3.3.7
EU Twinning Project BA 08-IB-FI-03
Institute for Statistics of Federation of BiH
Zelenih beretki 26, Sarajevo

Component 1 Business Statistics

- 1.1. Statistical Business Register
- 1.2. Structural Business Statistics
- 1.3. Construction Statistics
- 1.4. Retail Trade Statistics
- 1.5. Tourism Statistics

Component 2 Extended Household Budget Survey

Component 3 Institutional Capacity

- 3.1. Long term strategy
- 3.2. Metadata and classification system development

3.3. Promotion of dissemination

Benchmarks

- A) Staff trained in electronic dissemination by 8th project quarter
- B) Comprehensive training of the staff
- C) Planning for the building of a comprehensive dissemination database

3.3.7: PC-Axis training - population census

Background:

It has been decided to implement a dissemination database in Bosnia-Herzegovina using the PX-Web software. The software is developed by Statistics Sweden. An initial license has been procured by the twinning project and will cover the first year of implementation.

A first training mission on PX-Web (3.3.9 – Ellemose Jensen) and associated software was conducted in December (11-14) 2012. The general conclusion after this mission was that additional training would be appreciated and highly relevant. A new activity 3.3.10 PC-Axis staff training II was therefore added to the project and conducted by Ellemose Jensen at the end of January 2013.

The activities in December 2012 and again in January 2013 tested and explored various options for including housing and population census data in PX-Web and showed that this was indeed possible.

In 2013 a decision has been made to implement PX-Web / PC-Axis. In May 2013 an additional training activity was therefore requested by the Bosnian and Herzegovinian side. The decision to use PX-Web made the original planned activities for 3.3.7 obsolete and a side letter was therefore introduced to shift the focus of 3.3.7 to additional training in PC-Axis / PX-Edit / PX-Web.

Activities to be undertaken in preparation for the mission

- Prepare (10) PX-files one week before activity
- Prepare files for population statistics –
- Finish translation of user interface to local language(s)
- Decide on the number of languages that will be used in the dissemination of census results
- Discuss work share between statisticians, dissemination and IT
- Prepare list of variables and values that are expected to be used in the population census
- Prepare description of how data can be aggregated and supplied from the census organization
- Discussion / definition of how and where PX-Web is going to be installed in a production setup

Produced PX-files should be send by e-mail to the consultant no later than Thursday before the mission.

In the previous conducted missions the participants have bought lap tops with Microsoft Information Server installed. This has worked well and is also recommended for this activity. It is important that the participants have administrator right for the lap tops so additional software can be installed.

The expected activities are:

- Follow up and expansion of the pilot installations made in December and January 2012
- Continued training in PX-Web (on the job training)
- Continued training in PX-Edit for production of PX-files
- Work with multilingual tables (4 or more languages),
- Import of data from other files (File>Read),
- Split and join tables
- Web deployment process
- Test of the translated user interface
- Continued discussions on how population and housing census data can best be presented

Expected output

The purpose of the activity is to continue the work on implementing PX-Web in Bosnia-Herzegovina. The mission will allow the relevant stakeholders to produce the necessary PX-files to disseminate the results from the upcoming population census.

Time schedule

All times are tentative and subject to change

Time	Activity	Purpose	Required participants
Tuesday 9:00-10:30	Status of user interface translation	User interface of PX-Web must be translated to Bosnian / Serbian. A partial translation was started in December and should now be completed to a level where it can be systematically tested by users.	IT developers – (Agency?)
Tuesday 10.30-10.45	Coffee break		
Tuesday 10.45-12.00	Follow up on prepared files	Review of the files produced by the BiH side	
Tuesday 12.00-13.15	Lunch Break		
Tuesday 13:15-14:30	Follow up on prepared files	Quality insurance of work done – discussions of problems encountered	
Tuesday 14.30-14.45	Coffee break		
Tuesday 14.45-16.00	Follow up on prepared files		
Wednesday 09:00-10.30	Relevant tables	Discussion of tables that are relevant to the dissemination of census results	
Wednesday 10.30-10.45	Coffee break		
Wednesday 10.45-12.00	To be defined		
Wednesday 12.00-13.15	Lunch Break		
Wednesday 13:15-14.30	Preparation for census dissemination	Discussions on which files / content that will be required to disseminate population and housing census – Croatia and other Balkan countries as starting point	Staff assigned by BC for the project – IT / Dissemination
Wednesday 14.30-	Coffee break		

14.45			
Wednesday 14:45-16.0	Production of files for population census continued		
Thursday 09:00- 10:30	Production of files for population census – continued		Staff assigned by BC for the project – IT / Dissemination
Thursday 10.30-10.45	Coffee break		
Thursday 10.45-12.00	Production of files for population census – continued		
Thursday 12.00-13.15	Lunch Break		
Thursday 13.15 - 14.30			
Thursday 14.30 – 14.45	Coffee break		
Thursday 14.45 – 16.00	Continued		
Friday 09.00- 12.00	Final report – Signing final report		Staff assigned by BC for the project – IT / Dissemination

Annex 2. Persons met

Agency for statistics of BiH

Senad Mujkic, Professional advisor in dissemination sector
Sabrina Omerbasic, Professional advisor in dissemination sector
Adnan Gusic, IT
Nino Svjetlanovic, IT
Dzenan Kovacevic, Volunteer in dissemination sector

Institute for Statistics of Federation of BiH (FIS)

Amer Koric, Head of Department for database development, support to dissemination and data entry
Amel Sikiric, Technical solution advisor for data publishing in database

Institute for Statistics of Republika Srpska (RSIS)

Natasa Gojkovic, Senior statistician for confidentiality, data disclosure control
Igor Radan, Senior statistician for administration of statistical data bases

EUPHC2 – Technical assistance to Population and Housing Census Phase II

Mr Bozo Grbic – Expert in data dissemination
Mr Ede John – Team Leader EUPHC2

Twinning Project Administration 00 (387) 33/911-992

Bente Dyrberg, RTA, 00 (387) 66/657-967
Djemka Sahinpasic, RTA Assistant, 00 (387) 63/721-104
Haris Imamovic, Interpreter

Annex 3. Templates for PX-File creation

Templates in Excel for the creation of PX-Files with 2, 3 and 4 variables were prepared and distributed to the participants in electronic form during the activity.

Two variables:

	A	B	C	D	E
1	Popis 2013 BiH table 1 (2 variables)				
2	region	region	age	age	
3	reg1TOT	region1To tot	Total		1000
4	reg1	region1	0-19	0-19 years	300
5	reg1	region1	20-39	20-39 years	250
6	reg1	region1	40-59	40-59 years	250
7	reg1	region1	60-79	60-79 years	100
8	reg1	region1	80+	80 years and more	100
9	reg2TOT	region2To tot	Total		800
10	reg2	region2	0-19	0-19 years	200
11	reg2	region2	20-39	20-39 years	150
12	reg2	region2	40-59	40-59 years	100
13	reg2	region2	60-79	60-79 years	300
14	reg2	region2	80+	80 years and more	50
15	reg3TOT	region3To tot	Total		850
16	reg3	region3	0-19	0-19 years	200
17	reg3	region3	20-39	20-39 years	200
18	reg3	region3	40-59	40-59 years	300
19	reg3	region3	60-79	60-79 years	100
20	reg3	region3	80+	80 years and more	50

3 Variables

	A	B	C	D	E	F
1	Popis 2013 BiH table 2 (3 variables)					
2	region	region	age	age	civil	civil
3	reg1TOT	region1To tot	Total		sing	single
4					mar	married
5	reg1	region1	0-19	0-19 years	sing	single
6					mar	married
7	reg1	region1	20-39	20-39 years	sing	single
8					mar	married
9	reg1	region1	40-59	40-59 years	sing	single
10					mar	married
11	reg1	region1	60-79	60-79 years	sing	single
12					mar	married
13	reg1	region1	80+	80 years and more	sing	single
14					mar	married
15	reg2TOT	region2To tot	Total		sing	single
16					mar	married
17	reg2	region2	0-19	0-19 years	sing	single
18					mar	married
19	reg2	region2	20-39	20-39 years	sing	single
20					mar	married
21	reg2	region2	40-59	40-59 years	sing	single
22					mar	married
23	reg2	region2	60-79	60-79 years	sing	single
24					mar	married
25	reg2	region2	80+	80 years and more	sing	single
26					mar	married

4 Variables

	A	B	C	D	E	F	G	H
1	Popis 2013 BiH table 3(4 variables)							
2	region	region	age	age	civil	civil	sex	sex
3	reg1TOT	region1TOT	tot	Total	sing	single	m	male
4							f	female
5					mar	married	m	male
6							f	female
7	reg1	region1	0-19	0-19 year	sing	single	m	male
8							f	female
9					mar	married	m	male
10							f	female
11	reg1	region1	20-39	20-39 ye	sing	single	m	male
12							f	female
13					mar	married	m	male
14							f	female
15	reg1	region1	40-59	40-59 ye	sing	single	m	male
16							f	female
17					mar	married	m	male
18							f	female
19	reg1	region1	60-79	60-79 ye	sing	single	m	male
20							f	female
21					mar	married	m	male
22							f	female
23	reg1	region1	80+	80 years	sing	single	m	male
24							f	female
25					mar	married	m	male
26							f	female
27	reg2TOT	region2TOT	tot	Total	sing	single	m	male
28							f	female
29					mar	married	m	male
30							f	female
31	reg2	region2	0-19	0-19 year	sing	single	m	male
32							f	female
33					mar	married	m	male
34							f	female
35	reg2	region2	20-39	20-39 ye	sing	single	m	male
36							f	female
37					mar	married	m	male
38							f	female
39	reg2	region2	40-59	40-59 ye	sing	single	m	male
40							f	female
41					mar	married	m	male
42							f	female
43	reg2	region2	60-79	60-79 ye	sing	single	m	male
44							f	female
45					mar	married	m	male
46							f	female
47	reg2	region2	80+	80 years	sing	single	m	male
48							f	female
49					mar	married	m	male
50							f	female