

TWINNING CONTRACT

BA 15 IPA SR 01 17

Support to the reform of the statistics system in Bosnia and Herzegovina



MISSION REPORT

Component 3: Balance of Payment and IIPs

Activity 3.3: Training on transmission of data to Eurostat I

Mission carried out by
Peter Stoltze, Statistics Denmark
Laust Hvas Mortensen, Statistics Denmark

3 – 6 September 2018

Version: Final

Expert contact information

Peter Stoltze

Statistics Denmark

Copenhagen, Denmark

Tel: +45 3917 3827

Email: psl@dst.dk

Laust Hvas Mortensen

Statistics Denmark

Copenhagen, Denmark

Tel: +45 3917 3218

Email: lhm@dst.dk

Table of contents

- 1. General comments 4
- 2. Assessment and results 4
- 3. Conclusions and recommendations 5
- Annex 1. Terms of Reference 6
- Annex 2. Persons met 7
- Annex 3. Example R scripts 7

List of Abbreviations

BHAS	Agency for Statistics of Bosnia and Herzegovina
BiH	Bosnia and Herzegovina
CBBH	Central Bank of Bosnia and Herzegovina
EC	European Commission
EU	European Union
FBiH	Federation of Bosnia and Herzegovina
FIS	Institute for Statistics of Federation of Bosnia and Herzegovina
MS	EU Member State
RSIS	Institute for Statistics of Republika Srpska
RTA	Resident Twinning Adviser
ToR	Terms of Reference

1. General comments

This mission report was prepared within the EU Twinning Project ‘Support to the reform of the statistics system in Bosnia and Herzegovina’. It was the first mission to be devoted to ‘Training on transmission of data to Eurostat’ within Component 3 of the project.

The purposes of the mission were:

- Presentation of Statistics Denmark experience using R
- Preparation of transmission of data to Eurostat using R

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 their stay in Bosnia-Herzegovina and which highly facilitated the work of the consultants.

This 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, CBBH, Statistics Denmark, INSEE, Statistics Finland and Croatian Bureau of Statistics.

2. Assessment and results

CBBH has presented part of their production system for the IIP and BoP, which consists of a number of interrelated Excel spreadsheets. The system is very well organised, and the experts are of the impression that quality of the systems developed by the CBBH are excellent.

The experts have shared their general view on this way to organize production of statistical tables. As a general principle it is advised to keep data, calculations, and presentation (dissemination product) well separated. It is the experts’ assessment that the excel-based statistical production systems presented by CBBH has two drawbacks. First, the excel-format may not continue to be supported by the SDMX conversion tool, which can be solved by augmenting the existing statistical production system to make it more flexible in terms of output. Secondly, the excel-based production system makes the task of making relatively minor changes to the system time consuming, e.g. minor changes in classifications. This is because many tasks may have to be repeated and because substantial work has to be devoted at CBBH to ensure that no errors are incurred in the process.

The experts have presented two alternative tools (R and NADABAS) that may be of potential use in the work of CBBH. Both tools make it easier to track the process from input data to output, but with varying degrees of flexibility and cost. Both tools are open source and thus free of any license fees, but in both cases resources in the form of training of staff is needed.

Briefly, NADABAS shores up data in a database, which is used to consolidate data. R is a more general tool, which is used extensively in official statistics for a multitude of purposes. R can in principle handle the entire process from input data to the output, but the experts have focused on demonstrating how R can be used to augment the existing excel-based system.

The experts admit to have very little working knowledge on the topic of SDMX. Thus a lot of the mandatory results and benchmarks for the component as mentioned in the ToR have not been reached or fulfilled completely. As mentioned above, the experts believe that the alternative tools presented can provide a flexible and transparent basis for outputting datasets that will be accepted by the RDMX conversion tool.

The production of International Investment Positions (IIP) statistics was also presented by CBBH staff. One of the benchmarks of the component is the development of a methodological document describing the quarterly IIP statistics. It is the experts' assessment that the CBBH should be able to move to produce a methodological description of the statistical production system. In order for the experts to provide further input, it may be of particular importance to describe the sampling scheme, i.e. how businesses are selected to participate in the surveys. Also, the current methods used to handle non-response should also be described, e.g. how missing data from business surveys are replaced with court filings etc.

3. Conclusions and recommendations

The experts want to express their gratitude for being welcomed in a very friendly atmosphere as well as having their presentations and general ideas received in a very positive manner.

The experts do not hesitate to call the mission a success, as the staff involved in the introductory R training activity showed both great interest in and understanding of the potential usage of R.

- Both CBBH and the experts see potential use for R. This adheres both to the production of statistical tables and dissemination product (e.g. the SDMX report).
- Further training of the staff is needed before the current system can be supplemented or replaced with R based scripts and/or NADABAS.
- A business case has to be prepared before taking any binding decisions regarding replacing a currently functioning system.
- As a general principle it is advised to keep data, calculations, and presentation (dissemination product) well separated.
- The development of a draft methodological document describing the quarterly IIP statistics can be initiated.

Next activity is planned for October 8-12, and the content will depend on the forthcoming discussions within CBBH about the future use of R. If CBBH finds it worthwhile to continue the exploration of R, a number of targeted activities can be planned, e.g. specific production of a SDMX file. The proposed actions below are under the assumption that CBBH chose to do so.

Action	Deadline	Responsible person
Provide example of excel-files and libraries with very specific directions regarding input and output	Two weeks before next mission begin	CBBH
List with recommended R resources	One week before arrival in Sarajevo next mission	Laust/Peter
The development of a first draft methodological document describing the quarterly IIP statistics, specifically covering sampling and how non-response is handled	Before arrival in Sarajevo next mission	CBBH
Working example transforming Excel spreadsheet into flat file suitable as input for the SDMX Converter.	Before arrival in Sarajevo next mission	Laust/Peter

Annex 1. Terms of Reference**Terms of Reference****EU Twinning Project BA 15 IPA ST 01 17****Component 3: Balance of Payment and IIPs****3 – 6 September 2018****Activity 3.3: Training on transmission of data to Eurostat I****1. Mandatory results and benchmarks for the component**

Mandatory results:

- Quality criteria and quality reporting on balance payments statistics produced by 8th project quarter
- Quarterly international investment position statistics developed by 8th project quarter

Benchmarks:

- Quality criteria of balance of payment defined by 1st project quarter
- Draft quality report for balance of payment developed by 3rd project quarter
- Available data sources for quarterly international investment positions statistics identified by 4th project quarter
- Final quality report for balance of payment developed by 7th project quarter
- Methodological document for quarterly international investment positions statistics developed by 8th project quarter
- Consistency between quarterly and annual IIP performed by 8th project quarter
- First data on quarterly international investment position for selected years published by 8th project quarter

2. Purpose of the activity

- Presentation of Statistics Denmark experience using R
- Preparation of transmission of data to Eurostat using R
- Introduction in SDMX (sponsors for development, implementation supporter,...)
- Objectives of SDMX data exchange
- DSD requirements with particular insight in DSD related to BOP (including ITSS), IIP and FDI
- SDMX Global Registry (SGR)
- SDMX converting tool (requirements and use) – including practical exercise

- Input tables for the SDMX Converter - including practical exercise on creation of excel input tables
- Output tables and upload of data
- Data validation - including practical exercise on use of data validation tool
- Practical work

3. Expected output of the activity

- Preparation of transmission of data to Eurostat made
- Understanding and reporting by using SDMX
- Input provided to the ToR of next activity

Annex 2. Persons met

Central Bank of Bosnia and Herzegovina (CBBH)

- Amir Hadžiomerađić, Head of Statistics and Publications Department
- Vedran Milisav, Head of BoP statistics section
- Aida Kalčo, BOP Specialist
- Jelena Obradović, Senior BOP Specialist
- Maša Kamenica Jordamović, Senior BOP Specialist
- Naida Hadžimuratović, BOP Expert
- Ivana Gotovac, BOP Specialist
- Danica Lučić, BOP Expert
- Ervin Zolic, CBBH Expert

MS Experts

- Peter Stoltze
- Laust Hvas Mortensen

Twinning Project Administration

- Katja Møller Hjelvang, RTA
- Đemka Šahinpašić, RTA Assistant

Annex 3. Example R scripts

Script 1:

```
## where is data?
setwd("C:/data/dst/8 International Consulting/IR_Bosnien/R")

## this checks if needed packages are installed
packages <- c("tidyverse", "readxl", "openxlsx")
if (length(setdiff(packages, rownames(installed.packages()))) > 0) {
  install.packages(setdiff(packages, rownames(installed.packages())))
}

# Reading packages
library(readxl)
library(openxlsx)
library(tidyverse)

## reading data from MODULE 4 into excel
mod4 <- as.data.frame(read_excel("../data/4 - MAINTENANCE.xlsx", sheet =
1, skip = 3))

## a note of caution: if the columns contain a mix of types of data (e.g.
numeric and character) the data can become hard to work with
## you may need to convert variables from one type to another after
cleaning the rows

## cleaning up by removing rows and columns that are not needed
## keeping only rows from 4 to 876 and columns 1 to 4
mod4 <- mod4[1:873, 1:4]

## renaming variables
names(mod4) <- c("Code", "Type", "Country", "Value")

## converting mod4 to wide format ...and back to long again
## Convert to wide format
wide <- spread(data = mod4, key = Type, value = Value)

## note on the logic of the function 'spread()':
## in long format there are observations for some unit which is repeated
in the rows - in our case it is the various balances
## 'key' is the variable that defines how the values are different within
units - in our cases that is the variable Type

## this plots B, C and D against each other
plot(wide[, 3:5])

## this checks how many Bs are 0 (because 1+1 == 2 evaluates to TRUE,
which is the same as 1 rather than 0 if FALSE)
sum(wide$B == 0)
## ...and how many are NOT missing
sum(!is.na(wide$B))

## calculating the debit/credit-ratio
wide$ratio <- wide$D/wide$C

## convert back to long
long <- gather(data = wide, key = Type, ... = c(D,C,B,ratio))

long <- gather(data = wide, key = Type, value=amount, -c("Country",
"Code"))
long <- gather(data = wide, key = Type, value=amount, -(1:2))
```

```
## note on the logic of 'gather()': the key is now the name of a NEW
variable that contains information about where data in a single row in
the long data came from
## '... =' points to a list of variables that contains values to be
stacked in the long format - in our case C, D, B and the new variable
Ratio. If one variable is omitted, then the information is just dropped

## writing data to excel
write.xlsx(long, "mod4_new_version.xlsx")
```

Script 2:

```
# loading packages
library(readxl)

# set working directory as needed
setwd("C:/data/dst/8 International Consulting/IR_Bosnien/R")

# reading data
a <- read_excel("../data/3.b. Flat File sa Geo_5
Geo_6_2016_23_02_2018.xlsx", sheet=2)

# a is a tibble - don't worry about this
class(a)

# extract vectors with service codes and descriptions
k <- seq(from=4, to=187, by=2)

codeNumber <- unname(unlist(a[2,k]))
codeDesc <- unname(unlist(a[3,k]))
varName <- unname(unlist(a[4,k+1]))

# combine to new data frame
# can be done in several ways, e.g. data.frame
codeList <- cbind(codeNumber, varName, codeDesc)

# remove 3 first rows and columns with only variable names
a <- a[-c(1,2,3), -(k+1)]

# rename columns (including first three columns)
names(a) <- c("Code", "Transaction", "Country", varName)

# build list with Country-Countrycode correspondance
countries <- unique(a[,c(1,3)])

# finally convert variables to numerics
for (i in 4:95)
  a[,i] <- as.numeric(unname(unlist(a[,i])))

# You could go further:
# gather 291 countries * 92 services * 3 transaction types
# as precisely 80316 observations
# Also, verbose country name no longer needed
a_long <- gather (data=a, key="Service", value="Amount", -(1:3))
a_long$Country <- NULL

# Roughly 80 percent of observations are zeros...
sum(a_long$Amount == 0) / sum (!is.na(a_long$Amount))
```

```
# and save objects for further investigation
saveRDS (a, "FlatFile_sheet2_20180905.rds")
saveRDS (a_long, "FlatFile_sheet2_tidy_20180905.rds")
saveRDS (codeList, "codeList_20180905.rds")
saveRDS (countries, "countries_20180905.rds")
```