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 3.4: Training on transmission of data to Eurostat II

Mission carried out by

Peter Stoltze, Statistics Denmark

Laust Hvas Mortensen, Statistics Denmark

9 – 12 October 2018

Version: Final

STATISTICS DENMARK

Institut national de la statistique et des études économiques

Insee

Mesurer pour comprendre

Statistics Finland

CROATIAN BUREAU OF STATISTICS

Expert contact information

Peter Stoltze
Statistics Denmark
Copenhagen, Denmark
Email: psl@dst.dk

Laust Hvas Mortensen
Statistics Denmark
Copenhagen, Denmark
Email: lhm@dst.dk

Table of contents

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 second mission to be devoted to 'Training on transmission of data to Eurostat' within Component 3 of the project.

The purpose of the mission was:

- Further work on preparation of transmission of data to Eurostat.

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.

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

The following is a direct continuation of the assessment made during mission 3.3 for which the mission report is available here: https://dst.dk/ext/6669865443/0/BiH2018/3-3-Mission-Report-Training-on-transmission-of-data-to-Eurostat-I--pdf

The general advice is to separate the data from the presentation and dissemination products since the same data are used to produce several of these products. Binding the processing and presentation of the data to closely together makes the system very hard to maintain and not very flexible for addition of new extensions/modules.

To comply with this general advice the experts proposed building a modular workflow in R on top of the existing structure of the BoP data and consisting of the following modules:

1) Read data by period/country/type/transaction from Excel and store the data in 'tall' (and tidy) format.
2) Modify/transform the data as needed:
   a) Validate the individual data
   b) Form aggregates as needed
   c) Apply disclosure control
3) Write files for submission, either SDMX (i.e. XML) directly or flat file suitable for converter

This approach has a number of principal advantages:

- Step 1 allows the reuse of the existing structure, i.e. the 92 separate Excel documents typically prepared individually by a number of domain experts, does not have to be changed.
- Step 2 is where we can make different computations, e.g. verification that credit and debit are consistent with balance, or define new country groups for which aggregates should be added to the output.
- Step 3 gives the freedom to produce different output based on the same input.

During the mission a lot of practical training was devoted to becoming comfortable with the data wrangling verbs (functions for data manipulation) of the so called 'tidyverse' in R. Also the idea of organizing scrips in projects (a feature of RStudio) was introduced.

3. Conclusions and recommendations

We have made a setup (as a R project) which almost produces the desired SDMX files based on the 92 input files prepared by the CBBH. The setup was handed over to the CBBH staff, and they have been trained in making simple modifications to the setup.

A bit of refinement and documentation of the prepared setup is needed. Also, the last step of actually producing a valid XML output was not (entirely) achieved. The CBBH staff who will be working with R is recommended to participate in the European Statistical Training Programme of European Statistical System, which includes courses on the use of R in official statistics.

Actions needed for moving forward:

Action	**Deadline**	**Responsible person**
Verify input for system (content of the 92 basic components).		CBBH
Create rules for confidentiality flagging.		CBBH
Brush-up meeting via videoconference with key CBBH staff working with R	As needed (CBBH initiative)	Laust and 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
9 – 12 October 2018

Activity 3.4: Training on transmission of data to Eurostat II

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**
   - Further work on preparation of transmission of data to Eurostat

3. **Expected output of the activity**
   - Further work on preparation of transmission of data to Eurostat made
   - Input provided to the ToR of next activity

4. Participants

Central Bank of Bosnia and Herzegovina (CBBH)

- Amir Hadžiomeragić, Head of Statistics and Publications Department
- Vedran Milisav, Head of BoP Statistics Section

MS Experts

- Peter Stoltze, Statistics Denmark
- Laust Hvas Mortensen, Statistics Denmark

Twinning Project Administration

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

Agenda

Location: Sarajevo, Marsala Tita 25

Time	Day	Place	Event	Purpose / Details
9:00-11:00	09 October Tuesday	Hotel	MS experts meeting	Discuss the programme Preparation of activity
12:00-15:30	09 October Tuesday	CBBH	Meeting with BC Experts	
9:00-12:00	10 October Wednesday	CBBH	Meeting with BC Experts	
13:00-15:30	10 October Wednesday	CBBH	Meeting with BC Experts	
9:00-12:00	11 October Thursday	CBBH	Meeting with BC Experts	
13:00-15:30	11 October Thursday	CBBH	Meeting with BC Experts	
9:00-12:00	12 October Friday	CBBH	Meeting with BC Experts Debriefing with RTA	Presentation of MS experts' findings and agreement on the reached conclusions Agreement on mission report Identification of work to be done until next activity. Input to ToR of next activity Conclusions and next steps

Annex 2. Persons met

Central Bank of Bosnia and Herzegovina (CBBH)

- Amir Hadžiomeragić, Head of Statistics and Publications Department
- Vedran Milisav, Head of BoP statistics section

MS Experts

- Peter Stoltze
- Laust Hvas Mortensen

Twinning Project Administration

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

Annex 3: Scripts from BoP workflow

The following scripts were prepared during the mission (or shortly thereafter), and they shall not be considered production-grade code. Rather the scripts show the proposed process break-down into modules each serving a well-defined purpose. In the following listings the purpose of each script is described in the header of the script (which is always a good practice to exercise).

Note that the files are assumed to reside within a R project. Hence all files are referenced using relative paths.

MasterFile.R

```
# Masterfile with workflow for working with BoP
source ("R-functions/SourceFunctions.R")

# Import file locations and stuff
a <- ImportFileLocation("data/BasicComponents.xlsx")

# Import files
b1 <- ReadManyFiles(a[1:50,])

# Remove rows with country aggregates (identified by containing numbers!)
b1 <- b1 %>%
  filter (!grepl("[0-9]", code))

# Create list of integrety rules

# Apply integrety rules

# Create list of country groupings
g <- MakeGeoGroups("data/CountryGroupDefinitions.xlsx")

# Apply country gropings
b2 <- addTotals(b1, g)
saveRDS(b2, "data/eksempel.rds")
```

SourceFunctions.R

```
# This scripts sources a number of scripts and thereby
# making the functions defined available.
# It can be seen as a primitive version of a package...
# It also loads the required packages.

library(tidyverse)
library(readxl)

source("R-functions/ImportFileLocation.R")
source("R-functions/ReadManyFiles.R")
source("R-functions/MakeGeoGroups.R")
source("R-functions/AggregateCountries.R")
```

ImportFileLocation.R

```
# Import file and location

ImportFileLocation <- function(fname){
  if (file.exists(fname)){
    return (read_xlsx(fname))
  } else {
    cat("WARNING: The provided file does not exist!!!")
    return (NULL)
  }
}
```

ReadManyFiles.R

```
# Function for bulk reading of files...

ReadManyFiles <- function(a){
  # convert tibble to matrix
  b <- as.matrix(a)
  # get number of rows
  n <- dim(b)[1]

  for (i in 1:n){
    # build string with file path and name
    f <- paste("data/BoP/",b[i,1],".xlsx", sep="")

    # check that file exists (it should!!!)
    if (file.exists(f)){

      # read data from Excel
      item <- read_xlsx(f, sheet=b[i,4], range=b[i,5])
      # remove country
      item <- select(item, -Country)
      # adjust names
      names(item) <- c("code", "item", "amount")
      # add item description
      item$dim <- b[i,2]

      ## stack items
      if (exists("allItems")){
        allItems <- bind_rows(allItems, item)
      } else {
        allItems <- item
      }
    }
  }
  return (allItems)
}
```

MakeGeoGroups.R

```
MakeGeoGroups <- function (file){

  a <- read_xlsx(file, col_names = FALSE)

  nCou <- dim(a)[1]-2
  nAgg <- dim(a)[2]-2

  a1 <- as.matrix(a[3:(nCou+2),1:2])
  a2 <- as.matrix(a[1:2,3:(nAgg+2)])
  a3 <- as.matrix(a[3:(nCou+2), 3:(nAgg+2)])

  myGroups <- list()

  for (i in 1:nAgg){
    b1 <- unname(a2[1,i])
    b2 <- unname(a2[2,i])
    b3 <- a1[a3[,i]=="1",2]

    Group <- list(
      desc=b1,
      name=b2,
      def=b3
    )

    myGroups[[i]] <- Group
  }

  return (myGroups)
}
```

AggregateCountries.R

```
addTotals <- function(myTable, myGroupList) {
  newTable <- myTable
  # loop over aggregation groups
  for (i in 1:length(myGroupList)){
    # extract description and definition
    grp_desc <- myGroupList[[i]]$'name'
    grp_def <-  myGroupList[[i]]$'def'

    # calculate total
    total <- myTable %>%
      filter (code %in% grp_def) %>%
      group_by (item, dim) %>%
      summarise (amount=sum(amount)) %>%
      mutate(code=grp_desc, agg=T)

    # add total
    newTable <- bind_rows(newTable, total)
  }
  return (newTable)
}
```