

# **TWINNING CONTRACT**

**BA 12 IB ST 01**

## **Support to the State and Entity Statistical Institutions, phase VI**



## **MISSION REPORT**

**on**

### **2.2.4 Activity**

#### **Integrating business demography into SBS survey**

Mission carried out by  
Anne Katrine Bach Jensen, Statistics Denmark  
Jaakko Salmela, Statistics Finland  
Søren Netterstrøm, Statistics Denmark

26 - 30. September 2016.

Version: Final



Expert contact information

Anne Katrine Bach Jensen  
Statistics Denmark  
Sejrøgade 11  
DK-2100 Copenhagen Ø  
Denmark  
Phone:  
Email: [AKD@dst.dk](mailto:AKD@dst.dk)

Jaakko Salmela  
Statistics Finland  
Työpajankatu 13  
FI-00022 Helsinki  
Finland  
Tel: +358 229 3454  
Email: [Jaakko.salmela@stat.fi](mailto:Jaakko.salmela@stat.fi)

Søren Netterstrøm  
Statistics Denmark  
Sejrøgade 11  
DK-2100 Copenhagen Ø  
Denmark  
Phone:  
Email: [sne@viv.dk](mailto:sne@viv.dk)

**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

## **Executive summary**

### **1. General comments**

A lot of work has been invested by BHAS, FIS and RSIS in producing the Business Demography. Especially how to detect and deal with successions has been an area of focus. The key task to be done next is to write automatic programs which will automatize the matching according to name, location and activity as described in the manual to find potential successions that should not be treated as births or deaths. In addition, a database driven solution for the data storage should be implemented, MS Access or SQL server depending on the skills available.

The first draft of methodological document has been written. It describes the statistical offices' work process, which is in accordance with the OECD-Eurostat Manual on Business Demography. The document needs to be updated on certain topics, such as timing of each task, and detail for different matching cases. The document must be agreed upon by all parties.

### **2. Summary of the activities**

The overall situation of BD production in BiH was presented. The method to determine activity status of Enterprises from a BD view was demonstrated and found very useful. It is important to note that BD activity status may differ from administrative activity status in SBR.

Control questions in SBS survey regarding reorganizations were discussed. As inconsistency between the answers of these questions and information registered in the SBR were found, uncertainties of the results were expressed. It was discussed if the inconsistency were due to the fact that the survey questions were found difficult for respondents to understand and because received information may not match with the BD methodology and terminology. The MS experts advised not to compare data received from the survey with data registered in the SBR. The results were not found very usable by the MS experts.

The first draft of a methodological document on BD was presented and further discussed throughout the mission. The document provided a very useful starting point for further development.

MS Expert presented a proposal on how the work process could be organised. The document has been attached as Annex 3 to the report. This proposal was discussed in detail during Thursday and was later circulated for further comments before being included in the report.

MS Expert presented an updated version of BD production in Finland. This led to discussion on storing BD data separately from SBR data to avoid interference with existing data. Also, choosing enterprises from all NACE activities was discussed, as enterprises may change their economic activity over time and move in or out of BD scope over time. The data should be delineated only in the reporting step of the process.

Processing company and street names for matching was discussed. It was suggested to assign codes for all streets – this is a huge project but it will be a good investment for automating

future work. Improvement of this part of SBR will help the production of BD. The idea of creating a “thesaurus” of street names and abbreviations was presented by BiH as an interim solution. However it was not possible to estimate the resources needed for this approach and hence the feasibility of this approach. Another approach discussed was to ‘normalize’ street names through an automated procedure before making matches. This need further elaboration. Even if the current address data may cause automatic matching to produce insufficient results, even a partial solution is recommended. The BD quality will develop over time as the quality of underlying SBR data improves.

It was concluded, that beside the automated matching process regarding the detection of successions, if applicable, the manual verification should be performed on enterprises with more than 20 employees. It is not feasible to manually go through the entire population of possible new births and deaths.

Use of data on bankruptcies was briefly discussed. MS experts concluded that this data should be studied for possible use in SBR and from there also for BD.

## **Methodological issues**

In the view of the MS Experts sole proprietors (crafts) should be included in BD. The main source for these enterprises is the unique registration system as they do not submit financial statements. BiH is of the opinion that this requires further analysis. Crafts should be included in the Business Demography as they represent a large proportion of the business economy. As the purpose of BD is to facilitate economic policy analysis, they represent a key area of interest to data users in national and international spheres.

It was concluded that enterprises moving from one entity to another should not pose a problem for the BD production as the institutes exchange data on enterprises and there is only a limited number of such cases.

In the context of Business Demography, employment should be measured as the annual average over the operating period. This should be taken into account when using the employment figures from USR, especially when reporting results. For example, if an enterprise is in operation for the whole year, but has employees only for two quarters, the employment figure should be the average of four quarters. However, if enterprise operates for only half a year and has employees during both quarters, the employment should be average of the two quarters.

Storing the data on successions was briefly discussed. Each predecessor / successor could be recorded in the row for each birth / death. The final results can be calculated by combining annual tables. For survival by take-over –cases a single identifier should be chosen from the enterprises involved in the event.

## **What to do before the next mission for the BC Counterpart and Experts**

- Read all reports on previous mission to be summarized in the next and final mission.

## **Topics for the next mission**

- Summarize reports on previous missions and work done as well as results achieved

## **Conclusions and recommendations**

The overall methodological issues have been worked through and a useful methodological document has been written. This means that the methodology behind making the Business Demography has been established to the extent possible within the given framework, and is ready to receive input. Further work on the data input should now be the main focus in the further development of the Business Demography. This should include some sort of timeline for receiving and processing data input and deadlines for final output.

We recommend that the methodological document to be updated according to the different discussions we had during the mission. This should be done as soon as possible, such that no information is lost.

We further recommend that each entity implement, to the extent possible, and starts working with the document describing the work process producing the Business Demography, presented in Annex 3.

## Annex 1. Terms of Reference

### EU Twinning Project BA-12-IB-ST-01

#### Terms of Reference

**Component: 2.2.Structural Business Statistics**

**26.-30. September 2016**

**Institute for Statistics of FBiH**

**Sarajevo, Zelenih Beretki 26, first floor**

#### **2.2.4. Activity**

**Integrating business demography into SBS survey**

#### **Benchmarks**

- Methodology for Business Services developed by 8th project quarter
- Methodology for Business Demography developed by 8th project quarter
- New SBS variable on Business Services compiled and available by 8th project quarter
- New SBS variable on Business Demography compiled and available by 8th project quarter

#### **Purpose of activity**

The expected activities are:

- Presentation and discussion of carried out activities on business demography – Methodology and work tables;
- Presentation of Methodology for linking SBR with the compilation of SBS business demography by SBR
- Analysis of the responses of succession from SBS survey with birth and death enterprises from business demography
- Presentation of production of Business Demography in Finland and Denmark
- Problems in the production of business demography (coding of succession at the enterprise level, discussion about status of activity of enterprises that are in liquidation and bankruptcy in SBR)
- Discussion on customizing SBR IT application in order to improve procedures for business demography
- Discussion on data storage, maintenance and plans for regular updating

#### **Expected output**

- All problems discussed and given methodological solutions for production business demography
- Mission report – summarized and detailed version completed

## **Annex 2: List of participants**

### ***MS Experts***

Anne Katrine Jensen (DK),  
Jaakko Salmela (FI),  
Søren Netterstrøm (DK)

### ***Agency for Statistics of BiH***

#### Structural Business Statistics:

Tima Karačić, Head of Department for Structural Business Statistics,  
Bojana Cicović, Senior Official for Structural Business Statistics,

#### Statistical Business Register:

Dzenita Mustafić Specialist for SBR (Coordinator for SBR Component, BHAS)  
Mevlija Odočević Senior Advisor for SBR  
Ivana Tavra Čolo Senior Official for SBR  
Branislava Cvijetić Senior Advisor for IT  
Senija Fačić Head of Branch Office in Brčko District  
Vedad Osmanović Senior Advisor for IT in Brčko District

### ***Institute for Statistics of Federation of BiH (FIS)***

#### Structural Business Statistics:

Aida Ljuca, Senior Official for Structural Business Statistics,  
Alma Čolpa, Senior Official for Structural Business Statistics,

#### Statistical Business Register:

Enisa Rastić Head of Department for business, statistical register and classifications in FIS  
Enverea Hurić Senior Advisor for Economic classifications in FIS  
Razija Bičakčić Senior Advisor for IT in FIS  
Emira Beširević Senior Official for SBR in FIS

### ***Institute for Statistics of Republika Srpska (RSIS)***

#### Structural Business Statistics:

Danica Babić, Specialist for Structural Business Statistics  
Slađana Nikić, Specialist for Structural Business Statistics

#### Statistical Business Register:

Nataša Teinović Senior Official for SBR (video link)

### ***Twining Project Administration***

Søren Leth-Sørensen, RTA  
Djemka Sahinpašić, RTA Assistant  
Haris Imamović, Interpreter



## **Annex 3. Creating the basic data file for Business Demography from SBR data.**

During the mission this proposals was discussed and refined. It is based on the methodology presented by BHAS and introduces some extra steps in order to automate processing for all small units.

### **Step 1. Extract base data from SBR**

For the years x, x-1 and x-2 data for all Enterprises are extracted from SBR.

The following variables are extracted:

EnterpriseID

TAX-ID

Full\_name (of Enterprise)

Short\_name (of Enterprise)

Class of employment

Class of turnover

Responsible Entity

Entity of address

Settlement (code)

Settlement Name

Municipality (code)

Municipality Name

Streetname

NoHouse (House Number)

Main\_Activity\_code (and corresponding Sector code)

Legal Form (from Legal Unit)

Ownership (from Legal Unit)

Active\_inactive (from Enterprise)

Year from date\_of\_registration (Enterprise) = Establishment Year

Year from date from cessation date (Enterprise) (if closed)

Legal Form according to the needs Business Demography (rule needed)

BDActivity (to be calculated in Step 2, set to 0 (inactive).

BDBirth (to be calculated in Step 4, set to 0)

DBDeath (to be calculated in Step 4, set to 0)

This step is carries out by BHAS (SBR and IT) on request from Department of Structural Business Statistics.

This activity is carried out when data for Fiscal Year x has been received and processed by the entities and SBR following this has been exchanged between Entities and BHAS.

**Step 2. Determine statistically active/inactive**

According to the OECD-Eurostat Manual:

*The population of active enterprises including all employers and non-employers (N) consists of all enterprises that had either turnover or employment at any time during the reference period.*

BHAS is not able to carry out this step (except for District of Brčko) because the classes of Employment and Turnover does not distinguish between 0 and small values.

For each of the years  $x$ ,  $x-1$  and  $x-2$  the following table is produced

Enterprise ID  
Year  
BDActivity

BD activity is set to 1 if there is any row  
in Enterprise\_Financial\_Data for the Enterprise and Year with a value  $> 0$   
or in Enterprise\_Employment\_Data for the Enterprise and Year with a value  $> 0$   
(Note that both financial statements and other sources are used)

The tables only consists of those units that have BDActivity for the actual year.

The tables are produced for each entity and District of Brčko and delivered to BHAS.

The tables is then used to update the variable BDActivity in the tables produced in step 1.

**Step 3. Units in liquidation or bankruptcy**

Units in liquidation or bankruptcy may be considered closed even if they are not closed in SBR. If needed, cessationyear is set.

#### Step 4 Determine birth and death within year

Base on the data produced in step2 and 3 the variables BDBirth and BDDeath can be calculated.

Birth is calculated for year x and death is calculated related to year x-2.

For all units where cessation date is not null in data for year x, they are considered death in the year of the cessation date (administratively closed or closed due to bankruptcy / liquidation).

For other units the following schema is used

Enterprise Exist in x-1	Active In x	Active in x-1	Active in x-2	
N	Y	-	-	Birth
N	N	-	-	Registered but not active
Y	Y	N	N	New Active (Birth)
Y	Y	N	Y	Restarted in X
Y	Y	Y	-	Continuation
Y	N	Y	-	Dead (in year X-1)
Y	N	N	Y	Dead in Year x-2
Y	N	N	N	Inactive

If case of Birth (including New Active), variable BDBirth is set to 1 in data for year x,  
If case of Dead in year x-2, variable DBDeath is set to 1 in data for year x-2.

**In case you need to distinguish between Birth and New Active use codes 1 and 2 in BDBirth.**

It should be noted, that a unit may be born and die in the same year.

This step is carried out by BHAS.

**Step 4 Prepare for matching names and addresses.**

The next steps will use matching by name, location (entity, settlement, streetname) and activity code to eliminate 'new' units that should be seen as a continuation of an existing enterprise.

In SBR addresses are stored as entity, settlement (encoded), street name and house number. The issue has been raised if street names can be matched or not, due to potential misspelling and other differences.

The issue can be solved by encoding addresses (using a thesaurus). However, the total number of combinations of entity, settlement and street name in the table Enterprise is 53.224. In order to encode this all 53.224 would have to be manually checked and a street-code being assigned. If we assume one person could process 20 lines per minute, this would take approx. 45 hours. In the following years the thesaurus would have to be update, but this will be a much smaller exercise on dealing with new combinations (assumed to be a small number).

Another approach could be to do some normalisation of addresses. An experiment shows, that is everything is turned to upper case and period and spaces are removed then the number of combination can be reduced to 49.626<sup>1</sup>. If accents are ignored as well the number is reduced to 48.726.

The table Enterprise has in total 234.495 entries. Of these 188.120 is associated with a location with 3 or more occurrences (before normalisation). After normalisation this number increases to 191.809. This implies that a least 80 % of addresses are comparable. The cost of further 'normalization' that might slightly improve this number may not be worth the effort.

Names may be normalized in a similar matter.

It should be taken into account that there is an ongoing project to assign codes to street names (by municipality). If this is introduced in the donor registers for SBR it will solved the issue. However here is not any information about when this may happen. This may be further investigated.

This step is carried out by BHAS.

---

<sup>1</sup> SQL server: `SELECT .. REPLACE(REPLACE(UPPER(Streetname), '.', ''), ' ', '') collate SQL_Latin1_General_CP1_CI_AI AS NormalisedStreetName`

### Step 5 Automatic procedure to detect successions

According to the OECD-Eurostat Manual continuation is determined by matching units on Name, Location and Activity code. If two of these matches the new unit is seen as a continuation of the old unit. The Manual however does not further describe this matching. Do you need 3 or 4 digit Activity code, is Location settlement only, street or down to house number.

Units for year  $x$  with  $DBBirth = 1$  are compared to units in  $x-1$  on

- Location and Name
- Location and Activity code
- Name and Activity code

If there is 1 and only 1 matching unit in  $x-1$  the new unit in  $x$  is considered the continuation. For the new unit  $DBBirth$  is recalculated using data from the matching unit regarding year  $x-1$  and  $x-2$ . For the matching unit  $DBDeath$  is set to terminated in year  $x$ . (it is neither a birth or death).

In case of more than 1 match it is not possible to determine which old unit should be selected, these cases are ignored.

The automatic matching will not be accurate. Not all cases of continuation will be discovered and some of the cases marked as continuation will not reflect the real world. The first year this process is done, a small sample (50 or 100 units) should be taken out for manual inspection to determine to what extent the continuations found reflect real continuation. Data in SBR may be examined in more detail (using whatever sources are available).

A strong indicator of continuation may be that there is only 1 quarter where the old and new unit are paying social contributions (has employment).

In the following year (when  $x+1$  is available) a further analysis may be conducted by matching New units in year  $x$  (including those matched automatically) with Dead units in year  $x+1$ . If listed by settlement and activity code it may be easy to detect potential successions in order to assess the quality of the matching procedure.

There will be a need to experiment if location should be compared using house number or not and if activity code should be 3 or 4 digits. Perform analysis as above to determine the ratio of real matches from each approach to make the final decision. Once made it should be reused for at least a number of years.

If only street / 3 digit activity code is used, then cases with multiple matches could be re-examined using house number / 4 digit code to potentially finding a single match,

This exercise may be limited to units from Section A-N and division 95,96 for the year  $x$ . Large units will be covered in next step and may be excluded.

This step is carried out by BHAS. For manual inspection assistance from entities may be useful.

Matching by name may not give many hits but are included to follow the Eurostat-manual.

**Step 6 Manual procedure to detect successions**

As described above the automatic procedure will not be precise. For that reason units with more than 20 employees as well as a large turnover (threshold to be determined) should be handled in a manual process to determine potential succession. Only units within the frame.

All potential sources including taking contact to the enterprise should be utilized.

The units selected for manual inspection should be extracted to an excel worksheet.

It has to be determined if this task is carried out by BHAS or in cooperation with the entities.

**Step 7 Splits and Mergers**

Step 5 and 6 may be good in finding succession, however splits and merges may not be detected. The number of such event may not be large but however have some influence on the final results.

SBR contains some data regarding splits and mergers on the level of legal unit. This data, as well as other potential sources may be examined in order to detect such events and take appropriate actions in updating BDEvent for the units involved.

**Step 8 Create final register and sample frame.**

The result of the process carried out above is the final register for year  $x$  as well and registers for year  $x-1$  and  $x-2$  including DBBirth and DBDeath. The register for  $x$  and  $x-1$  should be used the following year as  $x-1$  and  $x-2$ .

In order to create tables as required by Eurostat, Number of Employed persons (average for year, period of activity) and turnover are needed. BHAS creates as file for each entity with units within the frame (A-N and 95,96) and with BDEvent as New, New active or Dead containing

Enterprise ID

Unified Enterprise (in case of succession, ID of first in succession line, else Enterprise ID)

Reference-year

DBBirth

BDDeath

Succession of (Enterprise ID)

Succeeded by (Enterprise ID)

BDActivity

ActivityCode

The entities then add number of employed persons and turnover and perform compilation of the figures needed. This steps need further to be developed in the methodology by BHAS.

Once figure are compiled by the entities they are sent to BHAS for final summing up, sending to Eurostat and other publishing.