



TWINNING CONTRACT

JO 21 ENI ST 01 22

Strengthening the capacity of Jordan's Department of Statistics in terms of compilation, analysis and reporting of statistical data in line with International and European best practices

MISSION REPORT

on

Component 3 **Strategy for missing or erroneous data**

Activity 3.2.2 **Sampling Frame**

Mission carried out by
Mr. Peter Tibert Stoltze
Dr. Andrius Čiginas
Mr. Tomas Rudys

Amman, Jordan
September 17-19, 2024

Draft version: 1.1

Authors' names, addresses, e-mails***Mr. Peter Tibert Stoltze****Head of division**Methodology and Analysis**Statistics Denmark**Sejrøgade 11**DK-2100 Copenhagen Ø**Denmark**Tel: +45 3917 3827**Email: psl@dst.dk****Dr. Andrius Čiginas****Specialist at Methodology and Data Science Group**Statistics Lithuania**Gedimino Ave. 29**01500 Vilnius**Lithuania**Tel: +37065668785**Email: andrius.ciginas@stat.gov.lt****Mr. Tomas Rudys****Head of Methodology and Data Science Group**Statistics Lithuania. State Data Agency**Gedimino Ave. 29**01500 Vilnius**Lithuania**Tel: +37068781037**Email: tomas.rudys@stat.gov.lt*

Table of Contents

Executive Summary	5
1. General comments	6
2. Assessment and results	6
3. Conclusions and recommendations	8
Annex 1. Terms of Reference (excerpt)	9
Annex 2: Programme for the mission	10
Annex 3. Persons met.....	11
Annex 4. Content of questionnaires	12

Strengthening the capacity of Jordan's Department of Statistics

List of Abbreviations

- BC – Beneficiary Country
- DoS – Department of Statistics
- MS – Member State
- PL – Project Leader
- RTA – Resident Twinning Advisor
- STE – Short-term Expert

- CoEE – Census of Economic Establishments
- CoP – Code of Practice
- CoPo – Census of Population
- SAE – Small Area Estimation
- SBR - Statistical Business Register
- ToR – Terms of Reference

Executive Summary

The Terms of Reference for this activity states: "A prerequisite for a successful mission will be access to all relevant data – Updated sampling frame, administrative data." As it turned out, neither the MS experts nor any of the BC participants had access to the required data. Hence, the planned program was abandoned, and alternative activities were conducted.

However, the MS experts had access to very small samples of the data. Based on this, a review of data formats, identifiers and possible use across data sources was given together with some comments on the outlined processes.

Also, the MS experts prepared a demonstration of how a frame update can be evaluated. The presented solution is based on Lithuanian data and R Markdown and is attached to this mission report.

1. General comments

This mission report was prepared within the Twinning Project "Strengthening the capacity of Jordan's Department of Statistics in terms of compilation, analysis and reporting of statistical data in line with International and European best practices". The current activity is carried out as part of component 3, Strategy for missing or erroneous data.

The actions planned for this activity were not carried out as scheduled due to one of the prerequisites (as per terms of reference) not being met. Specifically, access to relevant data was not provided: Neither to the experts, nor to any DoS experts participating in the activity and being able to conduct the needed exercises. This may or may not be due to misunderstandings during the planning of the activity (even though this is explicitly mentioned in the mutually agreed upon ToR, see last paragraph in Appendix 1). Since a thorough investigation of the updated sampling frame was the main purpose of the activity, the content of the activity changed accordingly.

The purpose of this activity was to review current situation in Jordan related to the Mandatory Results. The Twinning team, MS experts and DoS experts will obtain a common understanding of current situation, vision and challenges in relation to the Mandatory Results of this component. The level of compliance with relevant EU regulations and other international standards will be assessed.

The consultants would like to express their sincere thanks to all officials and individuals met for the kind support and valuable information which they received during the stay in Jordan and which highly facilitated their work. The views and observations stated in this report are those of the consultants and do not necessarily correspond to the views of the EU, Statistics Lithuania or Statistics Denmark.

2. Assessment and results

During the mission, we had access to very small samples of data from various sources. The accompanying spreadsheet (file: "Data and metadata overview.xlsx", sheet: "Source overview") lists these sources and data excerpts. The following is based on these sample data and various discussions with DoS staff and management.

The 2018 Census of Economic Establishments

The general Census of Economic Establishments (CoEE) was carried out in 2018. We have been presented with both the questionnaire (in English) and a small sample ($n = 248$) of the resulting data. The CoEE serves three purposes:

- General information about Jordanian enterprises as per 2018 (main economic activity, foreign trade, e-commerce, owner characteristics etc.);
- Targeted sampling frame for subsequent economic surveys at DoS (e.g. Industry Survey, Service Survey, Internal Trade Survey to name a few);
- Input to a planned Statistical Business Register (SBR).

A comprehensive document ("Methodology of The General Census of Economic Establishments", 22pp, available in English) describes some of the general information found through the 2018 CoEE as well as the applied methodology.

Strengthening the capacity of Jordan's Department of Statistics

Since establishments may change activity or altogether stop to exist, and since new establishments may emerge, the information in the CoEE will quickly be outdated. This means that the information will not be sufficient for use as a sampling frame choosing individual establishment for example with activities within the service sector. DoS has already acknowledged this and has conducted a subsequent update of the data (see below).

Finally, using the CoEE as input for a SBR is not straightforward because information in the CoEE is collected at the establishment level and the SBR is planned to cover the enterprise level. In many cases, an enterprise consists of only one establishment, and there is then a one-to-one (1:1) relation between the two unit types. But since an enterprise may also consist of two or more establishments (a many-to-one or n:1 relation), it has to be established how to (a) find out which establishments belong to the same enterprise, and (b) make rules for aggregating establishment level information to enterprise level information. Regarding (a), some of the questions in the questionnaire (especially the updated) tries to cover this aspect.

The 2023 Census of Economic Establishments update

In 2023, a survey targeted at updating the 2018 CoEE was established. The updating was done by selecting statistical blocks with either more than 35 establishments or significant individual establishments (where significant means that either number of employees or turnover is above a certain threshold). These blocks were then surveyed using a slightly modified version of the 2018 CoEE questionnaire. Appendix 4 present the basis for a review of the overlap between the two questionnaires.

Based on the 2023 survey, three possibilities for individual establishment exist:

- Establishment present in 2018 was not present in 2023 (“death”)
- Establishment present in 2018 was also present in 2023 (“survivor”)
- Establishment present in 2023 only (“birth”)

The consultants have seen neither data nor summary results of the 2023 update. Interesting facts would concern the number of births and deaths of establishments on block level, which with certain reasonable assumptions could have been generalized to the blocks that were not visited and hereby extent the update to the entire population. In addition, for the survivors, it would be interesting to know how they have evolved in the five-year period, e.g. describing the extent of change of activity.

We have been informed, that establishments present in 2018 but not in 2023 are not immediately removed from the data until further confirmation is found. This could lead to issues of over coverage. To some extent, it also defies the purpose of the update survey. The partial explanation for this process is the idea, that the establishment still exists albeit in another statistical block..

Planned 2025 update as part of the Census of Population

A Census of Population (CoPo) is scheduled for 2025. As part of this census, information about establishments is also gathered. To this end, section 5 of the questionnaire for the CoPo contains 10 questions concerning establishments (see Appendix 4).

Having interviewers in the field is obviously a good opportunity to collect (updated) information. However, for this approach to work, some clarifications about the sampling units have to be made. E.g. if the sampling unit in the CoPo is household, it is not straightforward

Strengthening the capacity of Jordan's Department of Statistics

to collect information about establishments. On the other hand, if the sampling unit is dwelling/building, then it might be possible. All in all, the connection between a social survey (population census) and a business survey is not clear.

3. Conclusions and recommendations

The updating of the frame is very valuable in order to minimize potential over- and under coverage in the economic surveys. However, if the updating is not done in a consistent way, it will lead to other frame imperfections.

Based on the samples of data presented, a few general advices on technical representation are in place:

- Split code and text and keep codelists/classification in separate files;
- Redundant variable should be omitted (e.g. code for geography is currently 14 variables but could be just one variable supplemented with a number of codelists, that could be shared across survys to foster horizontal coherence);
- Use date formats for variables containing dates instead of splitting into three variables (year, month, date).

It appears that the update of the 2018 data with 2023 data was done as a continuous process by editing directly in a "master" file. This is not a recommended approach, since it is neither possible to track the progress, nor to reproduce the process. As a minimum, a date variable for when the record was updated should be added. Also, it should be possible to reconstruct the file at different points in time, e.g. by the use of "frozen" versions.

As for now, the data sets that we were supposed to work with, do not have structural metadata, i.e. a systematic account of the content. Some information can be derived from the questionnaire used, but not all. As a minimum, the structural metadata should consist of:

- Variable name;
- Corresponding question in questionnaire (could be one-to-one);
- Valid values (e.g. integer, text, date);
- Representation (e.g. length of text field or codelist).

If this information is available, it is more straightforward to combine information across data sets, like updating one data set with information from another. Also, based on the information in the structural metadata, it is also possible to validate the data sets and make (simple) synthetic data sets, i.e. data sets with the correct structure but with non-sensitive content. Synthetic data sets allow external experts to make programs that could operate on the actual data without having access to them.

Finally, a number of various recommendations can be made:

- Include a variable indicating if the Establishment is from the formal or informal sector (could to some degree be inferred from whether or not the establishment has a national ID number or not);
- Move the frame from Excel to a database;
- When an address register and SBR are established, make sure that the frame can be linked to those.

Annex 1. Terms of Reference (excerpt)

1. Purpose of the activity

The purpose of this activity is:

- The Twinning team, MS experts and DoS experts will obtain a common understanding of current situation, vision and challenges for the updated frame for Economic Surveys in Jordan;
- Assessment of the level of compliance with relevant EU regulations and other international standards;
- Review and analyze the results from updated sampling frame for Economic surveys;
- Compare the results of a selected group of larger establishments against the administrative data;
- Assist with designing the documenting the updated sampling frame;

2. Expected output of the activity

- Mission Report;
- A common understanding of current situation, vision and challenges for the updated frame for Economic Surveys in Jordan obtained;
- The level of compliance with relevant EU regulations and other international standards assessed;
- The results from updated sampling frame for Economic surveys reviewed and analyzed;
- The results of a selected group of the larger establishment compared against the administrative data;
- Assist with designing the documenting the updated sampling frame;

A prerequisite for a successful mission will be access to all relevant data – Updated sampling frame, administrative data

Strengthening the capacity of Jordan's Department of Statistics

Annex 2: Programme for the mission

Programme according to the Terms of reference:

Day 1 (11:00 – 16:00)

- MS: Preparation in the Corp Hotel
- BC: National Holiday

Day 2 (09:30 – 15:00)

- BC: Presentation of the new frame
- MS and BC: Evaluation of the frame

Day 3 (09:30 – 15:00)

- MS and BC: Evaluation of the frame – continued
- MS and BC: Compare the results of a selected group of largest establishments against the administrative data

Day 4 (09:30 – 15:00) – Peer-to-peer work:

- MS and BC: Designing the documentation of the frame
- MS and BC: Summary of output and way forward

Annex 3. Persons met

DoS experts:

Directorate for Economic Statistics

- Mr. Mohammad AbelRazzaq, Director of Economic Statistics (DoS Component Leader), Mohammad.AbelRazzaq@DOS.GOV.JO
- Mazen Khalifeh - Head of Industry and Energy Division, mazen.khalifah@dos.gov.jo
- Faten Al-Ramini - Head of Services, Internal Trade, Financial and Insurance Division, faten.alramini@dos.gov.jo
- Mr. Yusuf Atharbeh, Head of Inputs and Outputs Unit, yusef.atharbeh@dos.gov.jo
- Mr. Mazen Khalifa , Head of Industry and Energy Unit, mazen.khalifah@dos.gov.jo
- Ms. Bailasan Sbaihat, Staff member of Directorate of Economic Statistics, bailasan@dos.gov.jo
- Mr. Mohammed Kheir Al-Amoush, Staff member of Directorate of Economic Statistics, mohdkhier.alomoush@dos.gov.jo
- Mr. Nasir Obeidat, Staff member of Directorate of Economic Statistics, nasir.obidat@dos.gov.jo

Directorate of National Account

- Mr. Murad Bani Hamad, Head of Annual Accounts
- Murad.BaniHamad@DOS.GOV.JO
- Ms. Wala'a Al-garam, Head of Quarterly Accounts
- walaa.gharram@dos.gov.jo




Twinning team:

- Eng. Tamer AL-Rousan, Director of Crop Statistics Division, Directorate of Agriculture and Environmental Statistics (RTA Counterpart)
- Ms. Charlotte Nielsen (RTA)
- Ms. Thekra Altorah (RTA Assistant)
- Ms. Zaina Amireh (Language Assistant)

Strengthening the capacity of Jordan’s Department of Statistics

Annex 4. Content of questionnaires

The following is an account of the questionnaires used for the 2018 CoEE, the 2023 update of CoEE, and the part of the 2025 CoPo concerned with establishments. In the table below is shown the first page of the questionnaires provided.

<p>2018 CoEE</p>	
<p>2023 CoEE update</p>	
<p>2025 CoPo</p>	

We have transformed these questionnaires into draft structural metadata in an accompanying spreadsheet. This makes it possible to investigate overlap in the collected variables and subsequently create suitable joining mechanisms between various datasets.