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 1
**Roadmap for the development of an integrated administrative data system in Jordan
with pilots on Statistical Business registers (SBR) and population statistics**

Activity: 1.2.7:
Quality assessment of administrative data for building a SBR

Mission carried out by
Mr. Ville-Matti Pilviö
Mr. Peter Tibert Stoltze

Amman, Jordan
13-16 May 2024

Version: Final

Authors' names, addresses, e-mails

Mr. Ville-Matti Pilviö
Senior Statistician
Statistics Finland
Työpajankatu 13, FI-00580 Helsinki, Finland
Tel: +358 29 551 3502
Email: ville-matti.pilvio@stat.fi

Mr. Peter Tibert Stoltze
Head of division
Methodology and Analysis
Statistics Denmark
Sankt Kjelds Plads 11
DK-2100 Copenhagen Ø
Denmark
Tel: +45 3917 3827
Email: psl@dst.dk

Table of contents

List of Abbreviations

- BC – Beneficiary Country
- CCD – Companies Control Department
- DoS – Department of Statistics
- MIT – Ministry of Industry and Trade
- MS – Member State
- PL – Project Leader
- RTA – Resident Twinning Advisor
- SBR – Statistical Business Registry
- SSC – Social Security Corporation
- STE – Short Term Expert

Executive Summary

Available sources of administrative data were reviewed during the mission with the aim to create register profiles for available administrative data sources. Administrative Data from Social Security Corporation (SSC) was used as pilot data. These were enhanced with Business Logic indicators that describe validations and quality checks needed at variable level for the data. The Business Logic was then used as input for creating code to implement the validation using the R-Progam and the validate package. The code was run on actual admin data to test its functionality. The SBR team should now be able to finalize the register profiles for the remaining administrative data sources and build the Business Logic for their variables.

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*". This Mission related to the following Mandatory Results (MR) and indicators:

***"MR 1.2: Pilot project to develop strategy for integrating administrative data sources for the purposes of creating an SBR**.*

- **Indicator 1.2.A:** Administrative data sources identified and assessed and plan developed for integrating these with Census of Establishments (CoE) information in an SBR

- **Indicator 1.2.B:** Documentation prepared on database structures and compliance with statistical standards, classifications (e.g. ISIC, Rev 4) etc. and use of common identifiers etc.

- **Indicator 1.2.C:** Explore how SBS can benefit other statistical domains in the DoS

The purpose of this activity was:
- To jointly review register profiles for available administrative data sources for building a SBR in Jordan – draft was provided before the mission;
- To establish a general framework for examining the quality of administrative data based on best international practices;
- To examine the quality of availible data sources for building a SBR in Jordan;
- To select variables for building a SBR in Jordan;
- To clean and standardise input data from administrative data sources;
- To complete structural metadata of data sources;

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 mission 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 EU, Statistics Finland or Statistics Denmark.

2. Assessment and results

What was achieved during this mission:
- Joint review of data and metadata available for building a SBR in Jordan;
- Documentation of admin data sources (creating and assess metadata);
- Methodology using Social Security Corporation (SSC) data as an example;
- Establishing a general framework for examine the quality of administrative data based on best international practices;
- Manual examination (understanding the process);
- Defining business logic (rules) at variable level using SSC data as an example;
- Using the programming language R and the „validate" package to implement the business logic;
- Reading and handling data with R (csv-files and spreadsheets with code lists and admin data);
- Confronting the data with the business logic, i.e. checking if data obeys the rules.

An annex to this mission report contains an example of how the validate-package in R can be used to check the data. During the mssion, real data (with Arabic texts) was used. In the provided example, an synthetic dataset is used instead.

Working with the register profiles is an important task that requires deep knowledge of the content of data and the processed involved in bringing the data to DoS. The register profile contains both general descriptions of the data source (reference metadata) and more detailed descriptions at the variable level (structural metadata). If this approach is adopted for all data sources, it will provide great value to DoS in terms of quality management.

3. Conclusions and recommendations

The Department of Statistics (DoS) SBR team is well on its way to setting up a working Statistical Business Register (SBR) with unit level data for both Legal Units and Local Units from Administrative Data sources. They have gained a strong understanding of available data sources and are in the process of finalizing register profiles for available administrative data sources. They have been introduced to a tool that will allow them the check the consistency between the data and the metadata. Asessing if the quality of a specific data source is sufficient to include this data source in the SBR is a task, that has only been briefly touched upon. Sufficient quality will always be a relative term dependent on the user needs, which needs to be established.

As StatBus version 1 was found lacking certain needed functionality's on testing at DoS, the SBR implementation awaits StatBus version 2 to be finalized after which the new version will be used to implement the SBR in Jordan.

Proposed next steps:

Action	Deadline	Responsible person
Complete metadata documentation for remaining data sources (in line with lessons learned)	By next mission	SBR team, Twinning team
Establish metadata management (i.e. avoid using locally stored spreadsheets with sprawling suffixes)	By end of next mission	SBR team, Twinning team, STE's
Complete business logic for validation of admin data variables for all data sources	By next mission	SBR team, Twinning team
Complete selection of best available variables to be used to build the SBR	By end of StatBus v 2 installation mission	SBR team, STE's
Installation of StatBus version 2 once a stable version is available	By end of StatBus v 2 installation mission	SBR team, STE's
Imputation of units at Legal Unit and Local Unit level into the SBR	2024	SBR team, STE's
Training and practice for the SBR team with R and PostgresSQL using DoS data and accessing their own databases	2024	DoS, Twinning team

Annex 1. Terms of Reference

Terms of Reference

EU Twinning Project JO 21 ENI ST 01 22

Component 1:
Roadmap for the development of an integrated administrative data system in Jordan with pilots on Statistical Business registers (SBR) and population statistics

Activity 1.2.7:
Quality assessment of administrative data for building a SBR

Dates: 13 – 16 May 2024

Content

List of abbreviations

BC   Beneficiary Country
DoS   Department of Statistics
ESS   European Statistical System
MS   Member State
RTA   Resident Twinning Advisor
STE   Short Term Expert
ToR   Term of References

1. Objective and Mandatory Results for the component
1.1 Objective
To prepare a roadmap for the development of an integrated administrative data system for Jordan, and conduct pilot projects on creating an SBR and strengthening population statistics.

As the development of a fully integrated administrative data system is a long-term project. The focus of the Twinning project will be on specific pilot projects where the use of administrative records can address key challenges currently faced by the DoS. These pilot projects will constitute the first steps in rolling out a roadmap for the Jordanian statistical system by providing a template for expanding the use of administrative data across the wider statistical system over time. The project will take outset in establishing the following core registers: (a) A Statistical Business Register (SBR); (b) A Population register and (c) A Dwelling and housing Register/Address register

The lack of a comprehensive SBR was identified as an obstacle to future improvements of DoS statistics back in the 2015 Twinning project led by Statistics Denmark. This gap still remains - developing and maintaining an SBR for Jordan would, therefore, greatly enhance the ability of the DoS to monitor and analyze developments in a rapidly changing business environment. The pilot project will examine available administrative data sources on business entities. In relation to business statistics, the development of a SBR it is essential to integrating different data sources, and providing additional analysis without increasing costs.

SBR is an essential component in supporting the coherence, consistency and quality of statistics produced as well as providing data on the population of business units. The SBR provides a basis for selecting and monitoring samples, integrating administrative and other data sources and providing a framework for deriving business demographics and National Accounts.

Recently the [Jordan Economic Modernization Vision 2030](#) was launched and "[Smart Jordan](#)" was identified as one of the eight Growth Drivers to implement the Economic Modernization Vision. The 'Smart Jordan Driver' includes seven sectors where data is one of them. This indicates the national interest to ensure constant and reliable data sources, and robust statistical systems that contribute to timely and informed policy making. It is expected that one of the measures that will be taken is to transform Jordan's Department of Statistics (DoS) into an interactive National Statistical Center (NSC).

1.2 Mandatory results and indicators for achievement for each sub-component

Component 1 is sub-divided in five sub-components each with a Mandatory Results (MR) and two to four indicators of achievements associated with the sub-component. In table 1 the two MR and related indicators of relevance for the current Mission listed in black.

Table 1: Mandatory results and indicators for achievement (IA) for each sub-components within Component 1: an integrated administrative data system for Jordan

MR from the Twinning Fiche	Indicator
MR 1.1: Compile an inventory of administrative data on business and households and an indicative roadmap for inclusion in an integrated system	**Indicator 1.1.A:** Inventory of administrative data variables and detailed supporting metadata prepared **Indicator 1.1.B:** Tentative roadmap prepared for inclusion of data in integrated system
MR 1.2: Pilot project to develop strategy for integrating administrative data sources for the purposes of creating an SBR	**Indicator 1.2.A:** Administrative data sources identified and assessed and plan developed for integrating these with Census of Establishments (CoE) information in an SBR **Indicator 1.2.B:** Documentation prepared on database structures and compliance with statistical standards, classifications (e.g. ISIC, Rev 4) etc. and use of common identifiers etc. **Indicator 1.2.C:** Explore how SBS can benefit other statistical domains in the DoS
MR 1.3: Undertake pilot project on how administrative records can be used to strengthen population statistics and inform framing of the 2025 CoP questionnaire	**Indicator 1.3.A:** Inventory of data sources prepared and assessed and action plan for incorporation in DoS statistics developed **Indicator 1.3.B:** Methodology developed for incorporating administrative data **Indicator 1.3.C:** Documentation prepared on statistical standards, classifications, identifiers, mapping etc. **Indicator 1.3.D:** Review of how administrative data can assist in developing the COP 2025 questionnaires
MR 1.4: Develop strategy for ensuring flows of data between the DoS and counterpart institutions are established on an ongoing basis for pilot projects above	**Indicator 1.4.A:** Review of technical infrastructure for data transfers and action plan prepared based on 1.1 and 1.2 above **Indicator 1.4.B:** MoUs agreed between DoS and partner institutions **Indicator 1.4.C:** Agreement on statistical standards, classifications, identifiers etc. between DoS and partner institutions **Indicator 1.4.D:** Review of data flows within the DoS
MR 1.5: Implement training programmes and develop training materials both within DoS and with partner institutions on the use of administrative records for statistical purposes, based on pilot projects above	**Indicator 1.5.A:** Detailed documentation on statistical standards, classifications, identifiers etc. developed. **Indicator 1.5.B:** Comprehensive training programs and workshops provided for DoS staff and partner institutions **Indicator 1.5.C:** DoS leadership role in ensuring proper statistical standards applied across the Jordanian statistical system reinforced.

2. Purpose of the activity

The purpose of this activity is:

- To jointly review register profiles for available administrative data sources for building a SBR in Jordan – draft was provided before the mission;
- To establish a general framework for examining the quality of administrative data based on best international practices;
- To examine the quality of availible data sources for building a SBR in Jordan;
- To select variables for building a SBR in Jordan;
- To clean and standardise input data from administrative data sources;
- To complete structural metadata of data sources.

3. Expected output of the activity

- Activity report;
- Register profiles reviewed;
- Indicators for quality of administrative data calculated for selected administrative sources – taking outset in input data for building a SBR in Jordan;
- Input variables for SBR selected;
- Methods and rules for standardizing SBR input data outlined;
- Structural metadata prepared for core variables;

4. Participants

4.1 MS Short Term Experts (STE's)

- **Mr. Ville-Matti Pilviö,** Senior Statistician, Information and Statistical Services, Data Resources, Statistics Finland. Mr. Ville-Matti Pilviö has more than 10 years' professional experience in official statistics administration. Currently, Mr. Ville-Matti Pilviö is responsible for The Business Statistics production System in Finland incl. Business, Register (SBR) used for the collection, analysis and production of Business Statistics. E-Mail: Ville-Matti.Pilvio@stat.fi

- **Mr. Peter Tibert Stoltze**, Head of the Division for Methodology and Analysis, Statistics Denmark.Mr. Stoltze has for more than 15 years worked in various positions with methodology issues at Statistics Denmark and coordinates methodology (sampling), metadata and quality work. Mr. Stoltze has extensive experience in the field of survey sampling and estimation, data editing, imputation and Quality assessment. In addition, Mr. Stoltze is experienced in providing support and training on methodological issues internationally. psl@dst.dk

4.2 DoS experts

COMPONENT LEADER AND HEAD OF THE SBR WORKING GROUP
Mr. Jaffar Ababneh, Director of Data Management Directorate
Jafaar.Ababneh@DOS.GOV.JO

DIRECTORATE OF DATA MANAGEMENT
Business Register Division
Ms. Safa'a Abu Ghaitah, Business Register Division in DoS- Staff/
safaa.aboeta@DOS.GOV.JO
Mr. Abd Alwahad Alharaizeh, Business Register Division in DoS- Staff /
Abdalwahed.ALharaizeh@DOS.GOV.JO
Ms. Anwar Al Khasbah, Business Register Division in DoS- Staff / anwarkh@dos.gov.jo

DIRECTORATE OF NATIONAL ACCOUNT
Mr. Murad Bani Hamad, Head of Annual Accounts
Murad.BaniHamad@DOS.GOV.JO

DIRECTORATE OF METHODOLOGIES AND DATA DIVISION
Quality Assurance Division
Mr. Derar Jawarneh, derar.jawarneh@DOS.GOV.JO

Participants for opening and closing sessions:
Dr. Tayseer Muqdadi, Technical Assistant to the Director General
Tayseer.Megdady@DOS.GOV.JO

The Twinning Team
Eng. Mohammad Khalaf, Director of Sustainable Development Unit, Department of Statistics
(DoS), Jordan (RTA Counterpart), Mohammed.Khalaf@DOS.GOV.JO
Dr. Charlotte Nielsen (RTA), cln@dst.dk
Ms. Zaina Amireh (Language Assistant), zainaamireh3@gmail.com

5. Tentative overall agenda

Day 1 (09:30 – 15:00):
- **MS:** Preparation for the Mission
- **BC:** Presentation of results from preparation actions taken by DoS before the Mission
- **BC and MS:** Joint review register profiles for available administrative data sources for building a SBR in Jordan – draft will be provided before the Mission

Day 2 (09:30 – 15:00):
- **BC and MS:** Cleaning and standardinzing input data from administrative data sources including structural metadata;
- **BC and MS:** Peer-to-peer activity on drafting general framework for examine the quality of administrative data based on best international practices.

Day 3 (09:30 – 15:00):
- **BC and MS:** Peer-to-peer activity on examining the quality of availible data sources for building a SBR in Jordan;
- **BC and MS:** Selection of variables to be used to build a SBR in Jordan based on the quality review
  NB: Prerequisite – access to data available

Day 4 (09:30 – 15:00):
- **BC and MS:** Follow up from the previous days
- **BC and MS:** Agreement of next steps to be taken
- **BC and MS:** Summing up and conclusion
- **MS:** Drafting the Mission Report

6. Background information for enhancing SBR at DoS

Data sources:

In brief, DoS has reach out to the following owner of administrative data.

1. The Ministry of Industry and Trade (MIT)
2. Companies Controller Department (CCD)
3. Social Security Corporation (SSC)
4. Income and Tax department
5. Greater Amman Municipality
6. Ministry of Local Administration
7. Ministry of digital Economy and Entrepreneurship

The current status is that the microdata for the following sources has been received and that metadata is currently being prepared and will be provided before the Mission

- The Ministry of Industry and Trade (MIT) (16 variables)
- The Companies Control Department (CCD) (10 variables)
- The Social Security Corporation (SSC) (28 variables)

Even though the Income and Tax Department has been identified as essential DoS has not been able to receive microdata data from Income and Tax Department. The obstacle for the lack of data from Income and Tax Department is that according to the law, the Income and Tax Department is only allowed to provide data to four institutions and DoS is not one of them.

Units to be in the future Statistical Business Register (SBR) in Jordan

Figure 1 illustrate the units and link between the different units to be included in a future Statistical Business Register (SBR) in Jordan when fully implemented. The model is based on United National International Guidelines for Development and Maintenance of Statistical Business Registers
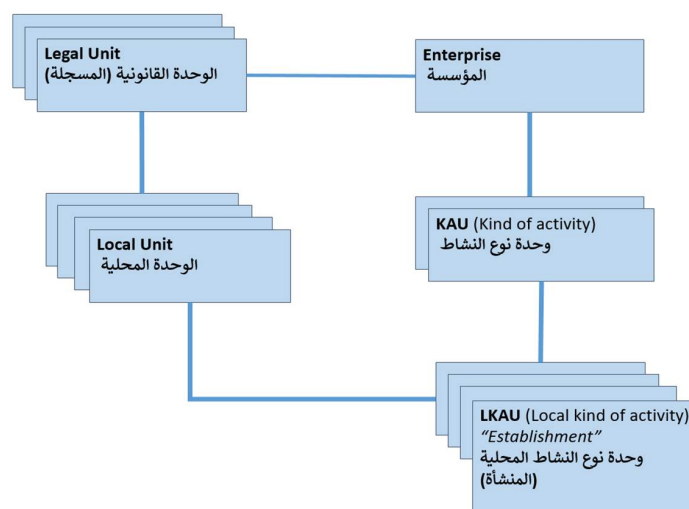(https://unstats.un.org/unsd/business-stat/SBR/Documents/UN_Guidelines_on_SBR.pdf)

Figure 1: *Conceptual model of units to be in the future Statistical Business Register (SBR) in Jordan*

StatBuS – The supporting IT system for building a SBR in Jordan
Currently, StatBuS is the supporting IT system being tested in Jordan. StatBus is the a client-server based system for statistical business registers which is an important backbone of any national statistical office. For additional information, please consult: https://www.StatBus.org/

The main features are:
- A tool for storing and maintain information on businesses from administrative sources, surveys and censuses
- Licensed as open source
- Based on UN International Guidelines For Development and Maintenance of Statistical Business Registers
- Used for drawing samples and make national economic statistics
- Important for National Accounts

The status for StatBus currently is that StatBuS has been successfully installed on a DoS server and test populations and key classification was uploaded to the system. Furthermore, it has been proved that the system support both Arabic and English. Mission Report from former Mission can be found on the project webpage: https://www.dst.dk/en/consulting/projects/jordan-2022-2024/component-1-integrated-administrative-data-system.

Detailed ideas for quality check of administrative data to be used for building a SPR in Jordan

Quality Check script (Validation) - What needs to be checked? :

List of validation checks on source AD files (Excel)

    Map source fields to SBR fields (documented)

    Import source AD file to validation environment (could be mssql or R)

    Number of columns = NCols /*Check, if (number of) AD variables corresponds (to number of) AD variables according to AD profile*/

    Number of rows = NRows /*Check, if number of AD units corresponds approximately to number of units expected*/

    List rows that have less cols than NColsand count()

    List rows that have more cols then NColsand count()

    List unique Ids in source that already exist inSBR - remove or overwrite? /*Dependent on which could be the causes ☐ Overwriting only if it is a more up-to-date information, otherwise research of causes. */

    List unique Ids in source that do not follow expected format of UIDs

    List any rows that have missing complusory fields, like UID, and the 5 other mandatory fields in Statbus

    Also include NULL and <blank> cells in the above check

    Which cols are numeric? List all rows that contain invalid (non-numeric data in cols that should be numeric and count()

    Which cols are varchar? List all rows that contain invalid (non-varchar data in cols that should be varchar and count()

    Which cols are dates? List all rows that contain invalid (date-data in cols that should be dates and count()

    Legal Forms: Which col is LF? Check this row for any invalid values (based on list of valid values) - List these rows and count() /*Coding list for legal forms or plain text?*/

    Activity Cat: Which col is AC? Check this row for any invalid values (based on list of valid values) - List these rows and count() /*Activity code according to ISIC?*/

    Other classifications that need to be validated?

    ➜ General check on double / triple datasets?

Importing excels into Statbus is quite frustrating at present. The above validations/checks aim to minimize problems with this.

This tool should be easy to update and modify for different AD sources. As new problems are identified, new checks should be added

For TCc SI, this might be carried out using MS SQL Stored Procs. The first priority is to complete this for Tax Authority AD. Once completed, AD from the Social Security will be validated with its own Stored Proc.

How to move forward with this?

a) Import data to R and carry out these validations?
Pros:
* R is a powerful tool for quick coding
Cons:
* SI skills to maintain R code?

b) Import data into mssql table and carry out these validations?
Pros:
* no new programming language skills needed
* support for easy excel importing
* easy access to existing classification and unit data (Statbus uses the same environment)
Cons:
* harder to code?

➔ Probably b as the Pros predominate?

Annex 2: Programme for the mission

Day 1 (09:30 – 15:00):
- o **MS:** Preparation for the Mission
- o **BC:** Welcome
- o **BC and MS:** Joint review of data and metadata available for building a SBR in Jordan

Day 2 (09:30 – 15:00):
- o **BC and MS:** Update of documentation of administrative data sources (creating and assess metadata);
- o **BC:** Methodology using Social Security Corporation (SSC) data as an example;
- o **BC and MS:** Manual examination (understanding the process);

Day 3 (09:30 – 15:00):
- o **BC and MS:** Peer-to-peer activity on examining the quality of availible data sources for building a SBR in Jordan
- o **BC and MS**: Selection of variables to be used to build a SBR in Jordan based on the quality review
- o **BC and MS:**Defining business logic (rules) at variable level using SSC data as an example;
- o **BC and MS:** Using the programming language R and the „validate" package to implement the business logic;
- o **BC and MS:**Reading and handling data with R (csv-files and spreadsheets with code lists and admin data);
- o **BC and MS:**Confronting the data with the business logic, i.e. checking if data obeys the rules.

Day 4 (09:30 – 15:00):
- o **BC and MS:** Follow up from the previous days
- o **BC and MS:** Agreement of next steps to be taken
- o **BC and MS:** Summing up and conclusion
- o **MS:** Drafting the Mission Report

Abbreviations:
MS = EU Member State (Denmark and Finland);
DoS = Department of Statistics, Jordan

Annex 3. Persons met

COMPONENT LEADER AND HEAD OF THE SBR WORKING GROUP
Mr. Jaffar Ababneh, Director of Data Management Directorate
Jafaar.Ababneh@DOS.GOV.JO

DIRECTORATE OF DATA MANAGEMENT
Business Register Division
Ms. Safa'a Abu Ghaitah, Business Register Division in DoS- Staff/
safaa.aboeta@DOS.GOV.JO
Mr. Abd Alwahad Alharaizeh, Business Register Division in DoS- Staff /
Abdalwahed.ALharaizeh@DOS.GOV.JO
Ms. Anwar Al Khasbah, Business Register Division in DoS- Staff / anwarkh@dos.gov.jo

DIRECTORATE OF NATIONAL ACCOUNT
Mr. Murad Bani Hamad, Head of Annual Accounts
Murad.BaniHamad@DOS.GOV.JO

DIRECTORATE OF METHODOLOGIES AND DATA DIVISION
Quality Assurance Division
Mr. Derar Jawarneh, derar.jawarneh@DOS.GOV.JO

Participants for opening and closing sessions:
Dr. Tayseer Muqdadi, Technical Assistant to the Director General
Tayseer.Megdady@DOS.GOV.JO

The Twinning Team
Eng. Mohammad Khalaf, Director of Sustainable Development Unit, Department of Statistics
(DoS), Jordan (RTA Counterpart), Mohammed.Khalaf@DOS.GOV.JO
Dr. Charlotte Nielsen (RTA), cln@dst.dk
Ms. Zaina Amireh (Language Assistant), zainaamireh3@gmail.com