**TWINNING CONTRACT**

**JO/13/ENP/ST/23**

# Strengthening the capabilities of the Department of Statistics in Jordan

# MISSION REPORT

**on**

**Activity 4.2: Plan for database structure - I**

Mission carried out by

Mr. Leonardo Tininini, Istat
Mr. Lars Knudsen, Statistics Denmark

2th to 6th March 2014

Version: Final

*Expert contact information*

*Mr. Lars Knudsen*
*Statistics Denmark*
*Sejrøgade 11*
*DK-2100 Copenhagen Ø*
*Denmark*
*Tel: +4539173318*
*Email: lak@dst.dk*

*Mr. Leonardo Tininini*
*Istat*
*Via Cesare Balbo 16*
*IT-00184Roma*
*Italy*
*Tel: +39 0646732504*
*Email:tininini@istat.it*

# Table of contents

# List of Abbreviations

| | |
|---|---|
| DoS | Department of Statistics of Jordan |
| Istat | Italian National Institute of Statistics |
| Dst | Statistics Denmark |
| ToR | Terms of Reference |
| SDS | Statistical Dissemination System |
| SDDB | Statistical Dissemination Database |
| CNMM | Common Nordic Meta Model |

# 1. General comments

This mission report was prepared within the Twinning Project "Strengthening the capabilities of the Department of Statistics in Jordan". It was the second mission to be devoted to planning database structures within Component 4 of the project.

The purposes of the mission were:
- To install the suggested Nordic Data model (CNMM)
- To test if the Nordic data model fits into and can work with DoS' other systems
- Techniques for ensuring a database which is homogeneous and flexible to meet future needs
- Definition of IT requirements for an optimal database structure which meets the need of DoS
- Plan for how to transfer data from old database to new database

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 Jordan and which highly facilitated the work of the consultants.

The views and observations stated in this report are those of the consultants and do not necessarily correspond to the views of EU, DoS, Statistics Denmark or Istat.

# 2. Assessment and results

The list of activities and discussions that took place during this mission is presented below in further detail. The main findings are summarized here according to the order of mission purposes above:

Between this and the previous mission, PX-Web was installed in DoS. It was verified that the installation works on px-files (the included test file from the installation). Furthermore, the CNMM 2.1 was also setup before this mission and a connection file (sqldb.config) was prepared. The installation should therefore be ready for data and metadata loading and DoS needs instructions on how to do this. During the mission the experts gave a walk through of the model, but concrete examples are crucial to the understanding of the model and its capabilities, and to ensure that the setup works as assumed.

Several discussions were devoted to the assessment of the database schemas used by DoS to store the surveys' microdata. In particular, DoS presented some general problems occurring in the survey on ICT usage (though can be generalized to many other surveys) and basically focussed on whether and how it may be possible to merge data from different rounds of the same survey and/or from different surveys. The experts illustrated that some positive results can only be achieved in very specific cases and require a tight cooperation between statisticians and IT staff. This is particularly true when considering sample surveys, where the possibility of data merging based on common keys and columns does not necessarily imply that such operation is meaningful (and hence acceptable) from the statistical point of view.

Provided that microdata merging is meaningful from the statistical point of view, the experts illustrated that a preliminary process of microdata reconciliation is required, which may greatly benefit from a centralized and harmonized repository of classification metadata. This classification repository would act as a reference for all microdata to be reconciled and specific mappings should be defined between microdata columns and classifications in the repository, to formally define their meaning and interrelationships. Instead of having one unique universal schema encompassing all different specific schemas, a more feasible approach should be based on a reconciliated (logical) "view" of the data, possibly limited to only a restricted number of classifications of interest (specifically those to be included in a given time series).

The experts did not discover any lack of hardware or software (technical requirements) that prevents the continuation of the process of reconciling data, or to continue the process of loading data into the Statistical Dissemination System (SDS). However, if the documentation on microdata structure is completely lacking and cannot be indirectly recovered in some other way (e.g. by a comparative analysis of the names of columns and identifying numbers on the paper questionnaire), microdata archival should be considered, leaving in the database only really useful (and practically usable) data.

DoS is using ESRI products for GIS, in particular in relation to the upcoming census. However, for dissemination purposes it is not clear what tool should be used. Thus, it was suggested by DoS that an overview of the old and new versions of PX-Map (for publishing maps on the internet) is given during the coming mission.

The activities of this and the next mission cover two main items: 1) The issue of micro data, and 2) how to disseminate data. To a large extent these items can be addressed in parallel. However, for this first mission concerning a "plan for database structure", DoS put high priority on the first item. Thus the second was only covered briefly. However, the second item is to be further addressed in the coming mission (4.3).

Overview of the activities and discussion:

**Day 1:**
Terminology overview
- The concepts of data warehouse, databases and statistical dissemination systems (SDS) was clarified

GIS discussion
- DoS presented the existing GIS systems and there was a discussion on ways of presenting GIS for dissemination purposes

Status on PX-Web and CNMM
- DoS showed the installation of PX-Web and the CNMM 2.1 (English) on Oracle

**Day 2:**
Discussion on merging data
- Whether or not data from different surveys can be unified was discussed, covering general aspects of sampling sizes and weights of the surveys

Discussion concerning "breaks" in data
- It was discussed how to handle different types of breaks in micro data registers, such as questions in questionnaires changing over time

CNMM
- The basic logic of the CNMM 2.1 was explained, using examples from www.statbank.dk

**Day 3:**
Data breaks in the SDS
- It was discussed how data breaks in the SDS can be handled, using examples from www.statbank.dk regarding changes in municipalities over time

Case study on ICT usage (1)
- DoS shortly presented three general problems occurring in the survey of ICT usage, which could be used as examples to look further into

Multiple schemas vs. one schema
- There were discussion on the pros and cons of using multiple schemas for a survey in contrast to using only one schema

Case study on Business surveys (1)

- Dos presented a new way of organizing data for business surveys, where several surveys are handled in one model. This model is different from models used in previous surveys

**Day 4:**
Case study on ICT usage (2)
- The problems mentioned in day 3 were further elaborated, and the experts gave possible solutions for the different types of problems, including how to deal with classifications changing over time, by reconciling data from e.g. different survey in one common view
Case study on Foreign Trade
- The existing foreign trade presentation system on [www.dos.jo](www.dos.jo) was presented. It was discussed how these data could be presented using PX-Web (e.g. with API functionality in PX-Web)
Case study on Business surveys (2)
- The experts gave some recommendations for changes concerning the model used for the Business surveys. The discussion involved problems with database 'hangs', which were addressed, by identifying some possible sources to be better investigated by DoS in the next weeks

**Day 5 (½ day):**
Conclusions and recommendations
- The experts presented their main conclusions from the assessments of the first four days, and also presented the derived recommendations
Agenda and further actions
- Based on this, the main topics to be addressed in mission 4.3 were agreed. A specific list of actions to be addressed before next mission was drafted (see below). Also the timing of next mission was discussed
Archiving micro data
- There was a discussion on how to setup a strategy for dealing with 'old data' and when to archive what

# 3. Conclusions and recommendations

**Conclusions**
Based on the assessments above, the following conclusions were drawn:
- Existing micro databases can be (at least partially) reconciled
- DoS needs a "Classification repository"
- DoS could greatly benefit from stronger internal cooperation (particularly between statisticians and IT/Dissemination)
- DoS have what is needed to use PX-Web and the CNMM
- DoS needs tools for map dissemination on the internet

**Recommendations**
Based on the conclusions, the recommendations are the following:
- Prepare case study data that can be reconciled with the purpose of dissemination in PX-Web
- A draft model for a common classification repository should be put in place
- DoS should consider how to improve internal cooperation between statisticians and IT/Dissemination staff
- Principles on when to archive register data should be drafted.
- Experts will shortly prepare a map road (with clear steps) to build a mapping and classification repository based on the case studies discussed during the mission.
- Dos staff should work on the steps defined in the road map and exchange information with the experts during the implementation.

**Proposal for agenda topics to be addressed during the next mission:**
- Review and refine the datamodel used for the case study "Employment and Unemployment"
- Prepare (reconcile) data from the case study for dissemination purposes
- Aggregate case study data for loading into the Statistical Dissemination System (SDS)
- Draft model for a classification repository
- Presentation of principles of GIS in SDS

*Actions needed for moving forward as well as for preparing the next mission:*

| Action | Deadline | Responsible person |
|---|---|---|
| Finalize powerpoints covering the discussions during mission 4.2 | Before next mission | Leonardo (Istat) |
| Prepare a test script for the CNMM | - | Lars (Dst) |
| Run and validate test script on DOS CNMM database | - | Hussam (DoS) |
| Choose classifications to be included in the reconciled view for dissemination (case study) | - | DoS in cooperation with experts |
| Prepare details regarding the classifications chosen according to expert indications | - | DoS |
| | | |

# Annex 1. Terms of Reference

---

## Terms of Reference

## EU Twinning Project JO/13/ENP/ST/23

## 2-6 March 2014

---

**Component 4: Data warehouse**

**Activity 4.2: Plan for database structure - I**

**0. Mandatory results and benchmarks for the component**
- New database structure defined and online dissemination improved (Apr 2015)
- Assessment report on current situation (Jan 2014)
- Develop a plan for the database structure (July 2014)
- Improve the IT-security (Jan 2015)
- Improve the online dissemination (Apr 2015))

**1. Purpose of the activity**
- To install the suggested Nordic Data model
- To test if the Nordic data model fits into and can work with DoS' other systems
- Techniques for ensuring a database which is homogeneous and flexible to meet future needs
- Definition of IT requirements for an optimal database structure which meets the need of DoS
- Plan for how to transfer data from old database to new database

**2. Expected output of the activity**
- Installation and thorough testing of the Common Nordic Metadata Model
- Proof of case that the Common Nordic Metadata Model can be used as the future database structure (data base for aggregated statistics)
- Decision on whether or not to use the Common Nordic Metadata Model as the future database structure (data base for aggregated statistics)
- Discussions of the general European Union experience with establishing micro data warehouse
- Discussion on how to treat the old data to include it into new database
- Discussion on how to treat the GIS data to include it into a new database
- Identify needs for training in relation to maintaining the database structure.
- Prepare recommendations on the future data base structure  inclusive of IT requirements
- Initiate work on plan for the transfer of data from the old to the new database structure
- Transfer of the Danish and in general the European Union, experience regarding database structure and online dissemination
- A lining up of work programme for the next activity (4.3)
- To decide upon the preferred timing of the next activity (4.3, Suggestion 20. April – 24. April?)

**3. Participants**

<u>DoS</u>
Mr Tayseer Deeb, Director of Information Technology (*Component Leader*)

Component team members…

<u>MS experts</u>
Mr Lars Knudsen, Chief Adviser, Web and Online Dissemination, Statistics Denmark
Mr. Leonardo Tininini, Head of Unit of Database Administrators, ISTAT

# Programme for the mission

| Time | Place | Event | Purpose / detail |
|------|-------|-------|------------------|
| Sunday, morning | Hotel / DoS | Meeting with RTA | To discuss the programme of the week |
| Sunday, morning | DoS | Meeting with BC Component Leader and BC Experts | |
| Sunday, morning | DoS | Meeting with BC Component Leader and BC Experts | |
| Monday, morning | DoS | Meeting with BC Component Leader and BC Experts | |
| Monday, afternoon | DoS | Meeting with BC Component Leader and BC Experts | |
| Tuesday, morning | DoS | Meeting with BC Component Leader and BC Experts | |
| Tuesday, afternoon | DoS | Meeting with BC Component Leader and BC Experts | |
| Wednesday, morning | DoS | Meeting with BC Component Leader and BC Experts | |
| Wednesday, afternoon | DoS | Meeting with BC Component Leader and BC Experts | |
| Thursday, morning | DoS | Meeting with BC Component Leader | Presentation of MS Experts' findings and agreement on the reached conclusions |
| Thursday, morning | DoS | Ad-hoc meetings | Final clarifications with BC Experts, preparation of report and presentation for BC Project Leader |
| Thursday, noon | DoS | Debriefing with BC Project Leader | Conclusions and decisions and their consequences for the next activity and the implied work programme for BC Experts |

# Annex 2. Persons met

<u>DoS:</u>

| Title/ Name | Directorate / Division | Email |
|---|---|---|
| Rana A. Swaidat | IT/ Head of Development Unit | rana@dos.gov.jo |
| Shatha Isa Muhammad | Programmer | Shatha.Muhammad@DOS.GOV.JO |
| Rania Abu Dhaim | Heaad of programming and analysis | Rania.AbuDhaim@DOS.GOV.JO |
| Tyaseer Anis | IT Director | tayseer.anis@DOS.GOV.JO |
| Ibrahim Ahmad | GIS Programmer | Ibrahim.AlKousheh@DOS.GOV.JO |
| Abdullah Al –sous | Web Dissemination section/IT | asous@dos.gov.jo |
| Manal Khuffash | Web Dissemination section/IT | manalk@dos.gov.jo |
| Mohammad Hasan | Network Engineer/ IT | Mohammad.Sakhrieh@DOS.GOV.JO |
| Hussam abu shukor | | hussam@dos.gov.jo |

Mr AbedWadood Matouk, Project Leader

Mr Ahmad Mowafi, Resident Twinning Adviser counterpart

<u>RTA Team:</u>

Thomas Olsen, RTA
Amal Aliah, RTA Assistant