

# **TWINNING CONTRACT**

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

## **Strengthening the capabilities of the Department of Statistics in Jordan**



### **MISSION REPORT**

**on**

#### **Activity 4.1: Assessment of current database structure**

Mission carried out by

Mrs. Annegrete Wulff, Statistics Denmark

Mrs. Lars Knudsen, Statistics Denmark

12<sup>th</sup> to 16<sup>th</sup> January 2014

Version: Final



Expert contact information

*Annegrete Wulff*  
*Statistics Denmark*  
*Sejrøgade 11*  
*DK-2100 Copenhagen Ø*  
*Denmark*  
*Tel: +45 3917 3155*  
*Email: awu@dst.dk*

*Lars Knudsen*  
*Statistics Denmark*  
*Sejrøgade 11*  
*DK-2100 Copenhagen Ø*  
*Denmark*  
*Tel: +45 3917 3318*  
*Email: lak@dst.dk*

## Table of contents

1. General comments.....	5
2. Assessment and results.....	5
3. Conclusions and recommendations .....	7
3.1 Overview of activities in component .....	8
Annex 1. Terms of Reference .....	10
Annex 2. Persons met.....	12

## **List of Abbreviations**

DoS	Department of Statistics of Jordan
ToR	Terms of Reference

## 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 first mission to be devoted to Assessment of current database structure Component 4: Data warehouse of the project.

The purposes of the mission were:

- Assessment and review on the current status of the current database structure, administration system as well as the current use of tools for online dissemination
- Priority to be given to the topics in the component
  - Database structure
  - Online dissemination
  - Website technology

The consultant would like to express his/her thanks to all officials and individuals met for the kind support and valuable information which he/she received during the stay in Jordan and which highly facilitated the work of the consultant.

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

## 2. Assessment and results

During the week discussions took place to let the experts have an overview of the activities:

1. Overview of current database structure
2. Overview of current administrative systems.
3. Overview of current tools used for online dissemination
4. Overview of current web site technology
5. Transfer of the Danish and in general the European Union, experience regarding database structure and online dissemination
6. Other issues to be addressed

Below the assessment results are presented.

### 1. Database structure

A large number of data requests are posed on the IT/Dissemination, e.g. concerning time series data. These requests consumes a lot of time, because data have to be retrieved from present and previous micro data (raw data), which are stored in different systems, and for older data not well documented.

The different types of systems for disseminating data were presented and discussed. The main types are:

- Surveys
- Census
- Sector specific (e.g. from external sources)
- Trade and Investment Information system (TIIS)
- JorInfo.
- GIS (however we did not manage to get it up on screen)

Each system has its own methods of selection and presentation for the end users. Some produce HTML while others return PDF (or shp from the GIS). In some cases the results are static tables and in other cases there is a dynamic selection process. However, the structure and formats of the resulting data tables are not flexible for the end users, as they cannot be flexibly filtered, modified or downloaded. Furthermore, data are structured so that it is not possible to make time series comparison.

Having many different systems to maintain and update is costly, and can make it difficult for end users to find data. It was mentioned that e.g. the TIIS system is reported difficult to use by end users.

Micro data are generally stored in Oracle and final data tables for end users are produced from a large number of Oracle procedures that specify the content of the final output tables.

Because of the different structure and contents of older surveys, in many cases it will be impossible to produce a coherent time series in detail on a specific statistical topic. And furthermore, it will be cumbersome and time consuming to produce data on an individual request basis. Thus, it was identified that a “hub with aggregated statistics” on the most detailed level possible, for each statistical domain, could be a solution. Such a system would, to a large extent, give both internal and external users the possibility to filter out the data they request by themselves, and download it in the format they require. Such a common data warehouse for aggregated statistics (macro data warehouse) could be build from existing data resided in Oracle. Newer data are available in a well defined structured form. For older data, the accessibility of data is depending on the structure of micro data and the level of documentation. However, when it concerns older data two types of challenges were addressed: 1) data from some previous surveys have got different questions, categories and structure. A direct comparison with the following survey is therefore not possible in the detail. Nevertheless these data can be disseminated as a time series if only the common data are shown, or presented at a higher aggregation level. 2) the second challenge is harder to overcome as it concerns surveys where the documentation is lost: There is no information on the content of the survey neither on the content of rows and columns. A solution could be to go to the printed publications of the old surveys, create the relevant tables in the database and enter the data manually.

Thus, the task can be divided into two separate items. One is to setup (design and implement) a common dissemination data warehouse for aggregated statistics (macro data), that have the ability to contain all different type of statistics on time series basis. The second item is concerned with populating the system with data. Which data can be loaded into the system depends on the availability of source data (micro- or macro data).

In the activity plan for this component, there is an item for presentation of dissemination tools, mentioning the PC-Axis software package. During this mission, data from the existing Oracle database (trade and employment data) were loaded into PX-Web for testing. It was concluded that the existing data seems to fit well into the data model/format for PC-Axis.

## **2. Administrative systems**

From the presentation of the administrative systems it became clear that these were irrelevant for the main project. They regarded staff, salary and other administration. Thus there will not be any more work around that in this component.

## **3. Online dissemination**

3. On the web site the external users have access to the published statistics in Arabic and English. It is presented according to the production, like *Surveys*, *Census*, *Jorinfo* and *Trade and Investment Information System*. It will be an improvement to have all statistical subjects in a common database to replace the current mix of HTML tables and pdf documents.

#### 4. Web site technology

Today the web site is build and maintained manually without a CMS. The content is to a large intent stored in Oracle and the presented tables created from Oracle procedures. A trend today among statistical offices – and in general - is to build the website in a CMS (content management system). In this way it is possible to separate the content from the design which make upgrading or design changes easier. And even decentralized maintenance can be carried out outside IT directly by subject matter staff. It also allows for making applications in CMS e.g. to secure a punctual update of the site.

#### 5. Transfer of experiences

It seems that no dissemination policy is in place to support the way content is presented on the web . Never the less, the IT staff study other statistical institutions website to get inspiration about best practice. Dissemination in Statistics Denmark and in Europe in general builds on the European Code of Practice for statistics.

([http://epp.eurostat.ec.europa.eu/portal/page/portal/product\\_details/publication?p\\_product\\_code=KS-32-11-955](http://epp.eurostat.ec.europa.eu/portal/page/portal/product_details/publication?p_product_code=KS-32-11-955)). Main principles on dissemination are *accessibility and clarity*. It is not sufficient that data are on the web site. External end users should be able to find them understand them and retrieve and use them correctly. This has to do with metadata. It is recommended that standardized structural metadata (lables and names for variables, values etc) are taken into account when creating tables for the internet site. Reference –or explanatory- metadata are handled in another component in the project-

#### 6. Other issues to be addressed (which are not included in the ToR)

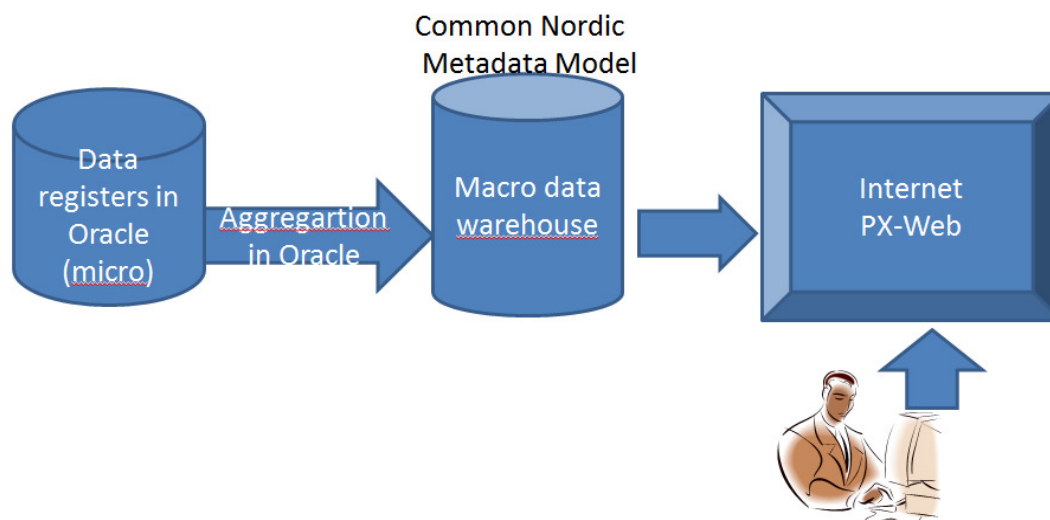
- There are certain problems when converting RDF data to other formats. It was requested if this could be an issue for coming missions.
- How to control sharp release of data for the web site
- Problem with Oracle that “hangs” in some cases, when several users update the same table at the same time.

There is a need for general Oracle DBA training (expertise in DBA training to analyze and build databases (theoretical and practical) before the missions on implementing a data model is carried out.. In general it was the impression of the experts that relevant capacity within the Dos on Oracle, C#, HTML, JavaScript etc.exists.

### 3. Conclusions and recommendations

The following conclusions and recommendations are referring to the numbering in section 2 above.

1. Having several different dissemination systems for different types of statistics is costly. External users will expect to find common ways of finding, processing and downloading statistics across different statistical domains. Thus, a common data warehouse for aggregated statistics seems to be the best solution.



PC-Axis software was tested and found compatible with existing data.

Before taking the final decision on implementing a common data warehouse for aggregated statistics, which can be used also as the basis for the online dissemination, it should be thoroughly tested, if the suggested Common Nordic Metadata Model, can function together with DoS's other IT-systems. Also, financial support from DoS's budget for the payment of licences after the Twinning project, should be secured.

2. This item turned out not to be relevant for the component.
3. It is recommended that a formalized cooperation between the subject matter units and the IT/Dissemination is being established in order to define responsibilities.
4. It is recommended that DoS in connection within the process of designing a new web site takes into account the possibilities in saving resources and getting easier maintenance of the website, contemplates the possibility of building the new site in a CMS.
5. It is recommended that DoS studies the Code of Practice to get inspiration on best practice, principles 13 (*Timeliness and punctuality*), 14 (*Coherence and Comparability*) and 15 (*Accessibility and Clarity*) in dissemination. Regarding the metadata it is recommended that participants from this component 4 can participate partly where relevant in the metadata component.
6. It has to be determined whether the issues mentioned above are to be included in the following mission, and if so, in which activities. It has to be identified if this has any effect on the planning of the coming activities.

### 3.1 Overview of activities in component 4

Component 4 Data warehouse covers different aspects of disseminating aggregated statistics to the public. Activity 4.2 and 4.3 will work around defining the structure for a common data warehouse containing aggregated non-confidential data. The assessment mission 4.1 concluded a possible solution could be a trial installation of the Nordic metadata model. Activity 4.4 will focus on security issues on the basic infrastructure needed to get online with the database. When it has been dedicated to work with the Nordic database model and pc-axis activities 4.6, 4.7, and 4.8 will be dedicated to setting up a road map for data to be entered into the px-web database, training in using the tools etc.



Activity 4.9 and 4.10 will cover the web site and how to improve the usability and maintenance work. The final mission 4.11 will review the results of the component.

<b>Action</b>	<b>Deadline</b>	<b>Responsible person</b>
Install the data model ( Nordic metadata model)	Up till next mission 4.2	DoS, Dissemination
Set up a prioritized plan for entering data to the new data model	Up till next mission 4.2	DoS
Investigate where to find training competence in DBA (Oracle)	Up till next mission 4.2	DST/RTA
Download and study how to implement the PC-Axis package <a href="http://www.scb.se/pc-axis">www.scb.se/pc-axis</a>	Up till next mission or before mission 4.6	DoS
Decision on using the Pc-Axis as for dissemination, incl the datamodel (the project pays year 1 and 2 hereafter pays DoS 5000 SEK per year = approx. 730 JD)	Before 1 July 2014	DoS
Establish a working cooperation with partners in statistical subject divisions for the relevant prioritized data to be entered	Before mission 4.6	DoS
Study the European Code of Practise , principles 13, 14, 15 and make a plan for how to use this in DoS web site dissemination	Before mission 4.9	DoS

## Annex 1. Terms of Reference

<p style="text-align: center;"><b>Terms of Reference</b></p> <p style="text-align: center;"><b>EU Twinning Project JO/13/ENP/ST/23</b></p> <p style="text-align: center;"><b>12-16 January 2014</b></p>
---

### Component 4: Data warehouse

#### Activity 4.1: Assessment of current database structure

#### 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

- Assessment and review on the current status of the current database structure, administration system as well as the current use of tools for online dissemination
- Priority to be given to the topics in the component
  - Database structure
  - Online dissemination
  - Website technology

#### 2. Expected output of the activity

- Overview of current database structure
- Overview of current administrative systems
- Overview of current tools used for online dissemination
- Overview of current website technology
- 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 (4.2, scheduled to 2nd - 6th March 2014)

#### 3. Participants

##### DoS

Mr Tayseer Deeb, Director of Information Technology (*Component Leader*)

Rania Abo Dhiem, IT

Hussam Abu Shukor, Dissemination

Rana Sweidat, IT

Mohammad Sakhrieh, IT

##### MS experts

Mrs Annegrete Wulff, Head of Division, Web and Online Dissemination, Statistics Denmark

Mr Lars Knudsen, Chief Adviser, Web and Online Dissemination, Statistics Denmark

## Programme for the mission

<b>Time</b>	<b>Place</b>	<b>Event</b>	<b>Purpose / detail</b>
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	Briefing of Danish data warehouse and server's specification (30 min) Discussion of DoS current database structure
Sunday, morning	DoS	Meeting with BC Component Leader and BC Experts	Discussion of DoS current database structure
Monday, morning	DoS	Meeting with BC Component Leader and BC Experts	Discussion about compatibility problems between old and current data and standardization
Monday, afternoon	DoS	Meeting with BC Component Leader and BC Experts	Presentation of how databases are built at Statistics Denmark
Tuesday, morning	DoS	Meeting with BC Component Leader and BC Experts	Discussions of the status regarding how to publish data from DoS on the website. Discussion of the fundamentals in building website and publishing data
Tuesday, afternoon	DoS	Meeting with BC Component Leader and BC Experts	Presentation of technology and fundamentals used in for Statistics Denmark's online dissemination
Wednesday, morning	DoS	Meeting with BC Component Leader and BC Experts	Discussions of how to document the data in each of the website and database (metadata)
Wednesday, afternoon	DoS	Meeting with BC Component Leader and BC Experts	Discussions of how to document the data in each of the website and database (metadata)
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**

### DoS:

Rania Abo Dhiem, IT

Hussam Abu Shukor, Dissemination

Rana Sweidat, IT

Mohammad Sakhrieh, IT

### RTA Team:

Amal Aliah, RTA Assistant

Deena Moghrabi, Interpreter

Thomas Olsen, RTA