



**EU Twinning Project  
IS12/ENP-APFI/o8**

**Support to the Israeli Central Bureau of Statistics  
in the development of National Accounts, Education Statistics,  
Survey Methodology, ICBS Website and  
Coordination of Israel National Statistical System**

**Component E  
Dissemination and Website**

**Activity E6.1  
Integration with DataWareHouse  
of dissemination and website**

***Implemented by:***

Lars Knudsen  
Chief Adviser  
Dissemination  
Statistics Denmark

**Jerusalem**  
20-23 October 2014



**Author's name, address, e-mail**

*Lars Knudsen  
Statistics Denmark  
Sejrøgade 11  
2100 Copenhagen Ø  
Denmark  
Tel. +45 3917.3318  
[lak@dst.dk](mailto:lak@dst.dk)*

## Table of contents

1. General comments.....	4
2. Assessment and results .....	4
<i>Principles of DWH building</i>	
<i>Data from existing data sources to the DWH</i>	
<i>Cube design and mini-pilot</i>	
<i>Updating rules and procedures</i>	
<i>Integration with web site</i>	
<i>Linking to quality metadata</i>	
<i>Work plan for developing the DWH</i>	
3. Conclusions and recommendations .....	6
4. Roadmap.....	6

## Annexes in the report

Annex E6.1.1 Terms of Reference.....	14
Annex E6.1.2 Meeting program .....	15
Annex E6.1.3 Persons met .....	16
Annex E6.1.4 Mini-Pilot with ICBS data .....	9
Annex E6.1.5 Relevant links .....	12

## Annexes external to the report (on the project's website: [www.dst.dk/israel](http://www.dst.dk/israel))

Annex E6.1.6 Intro to the dissemination database in Statistics Denmark (Power Point)	
Annex E6.1.7 API and mobile access to StatBank Denmark (Power Point)	
Annex E6.1.8 New data cube in the statistical dissemination system (Power Point)	
Annex E6.1.9 PC-Axis software family overview (Power Point)	
Annex E6.1.10 Dissemination Database and its use (Power Point)	
Annex E6.1.11 Debriefing E6.1 (Power Point)	

## List of Abbreviations

API	Application Programming Interface
BC	Beneficiary Country (Israel)
BoI	Bank of Israel
CBS	Central Bureau of Statistics (Israel)
CMS	Content Management System
CNMM	Common Nordic Metadata Model
CoP	Code of Practice
DDI	Data Dissemination Initiative
DWH	DataWareHouse (for aggregated statistics)
MS	Member State
NSI	National Statistical Institute
QAF	Quality Assurance Framework
UX	User Experience

## **1. General comments**

This activity was the seventh in component E -Dissemination and Website. It was implemented according to the Terms of Reference and the Agenda, cf. Annexes E6.1.1 and E6.1.2

The MS Expert would like to thank the Israeli side at the CBS for their strong commitment to the project and for their open-mindedness resulting in both interesting and challenging discussions.

The report does not necessarily reflect the views of Statistics Denmark or the EU, but are solely the MS Expert's views and recommendations.

## **2. Assessment and results**

According to the Terms of Reference and the agenda of the mission, the following items were dealt with during the meetings:

- Principles of DataWareHouse (DWH) building
- Getting data from existing data sources to the DWH
- Cube design for specific statistical subject
- Updating rules and procedures for the DWH
- Integration between web site and DWH
- Linking quality metadata to the DWH
- Work plan for developing the DWH

Each of the items is explained in further detail below.

### **Principles of DWH building**

This activity was divided into the following two items:

- Discussion of the current status in ICBS and in Statistics Denmark
- Introduction to the Common Nordic Metadata Model (CNMM)

ICBS pointed out that their strategy is to have one common DWH to serve all dissemination channels. This is also the strategy of Statistics Denmark, and was referred to as “the single source principle”.

Based on the information obtained during the first point, it was proposed to use the CNMM (see annex E6.1.5 for links and further information) as the relational data model for building a DWH for aggregated statistics.

This model supports the creation of “multi-dimensional data cubes” that can be used to “slice and dice” data, i.e. dynamically letting a user or a system select a subset of a large data cube, for specific data presentation e.g. as a table or graph. By using an Application Programming Interface (API) on top of this model, it will be possible to achieve the goal of having one DWH that serves all dissemination channels.

### **Data from existing data sources to the DWH**

The main steps of getting metadata and data into the DWH were presented, cf. Annex E6.1.x. At ICBS, a number of relevant data sources exist (such as the Informix, SQL-Server, Oracle, Excel etc.). To populate the DWH with metadata, it is important to use existing sources as well. For this the ICBS developed “Dictionary” could be used. It would be possible to retrieve data from other metadata sources as well, e.g. Colectica.

It was a wish from the ICBS side to use the “Informix Time Series Database” for a mini-pilot, to see how these data would fit into the DWH. During the activity, focus was on getting data into the DWH in two different ways 1) using a database table/view, and 2) using a CSV file. It should be possible to produce one of these outputs from any of the ICBS data sources.

However, using the “Informix Time Series data” as input for the DWH proved to have limitations. The main point is that time series are not multi-dimensional, and thus do not provide the detailed information that is preferred when designing multi-dimensional cubes. Thus, even though data can be put in the DWH, the advantages of cubes in a DWH, is not necessarily reached. This issue is further addressed in the examples below.

### **Cube design and mini-pilot**

The MS expert showed how foreign trade data from the ICBS web site can be put into the DWH as a data cube. Then, the CNMM version 2.1 was installed on ICBS MS SQL-server. A cube with educational statistics was designed and created using time series data from Informix. The result of this exercise is presented in Annex E6.1.4 below.

### **Updating rules and procedures**

The work flow and procedures of Statistics Denmark was presented, and it was discussed how the optimal flow in ICBS could be obtained. This resulted in a work plan (or road map) which is described in section 4.

### **Integration with web site**

The DWH must have an API functionality for it to be useful in different dissemination channels. In the web site, the API can be used to retrieve already updated data for tables, graphs, maps etc. The API can also be used to create an application that exposes all the data of the DWH in different ways. As inspiration, the mobile version of StatBank Denmark was presented (m.statbank.dk). It was also discussed how the API can be used to create dynamic web pages and publications.

### **Linking to quality metadata**

As mentioned above, the DWH should retrieve its metadata from existing metadata sources (such as the dictionary system). But also descriptive metadata (or quality metadata) can be linked to the cubes in the DWH. The CNMM does not contain an internal model for quality metadata, but tables can be linked to any kind of externally placed quality reports (in PDF, web sites etc.).

It should be a strategic goal that all data in the DWH have some kind of quality reports attached, for the end users to have the relevant documentation when they use the data.

### **Work plan for developing the DWH**

During the E6.1 activity, the main elements to achieve a successful DWH were identified. These items are:

- Tools for inserting and updating data and metadata in the DWH
- Source data (Informix, SQL-Server, Oracle, CSV etc.)
- Source metadata (Dictionary etc. for building cubes and quality metadata for end users)
- Guidelines for designing data cubes, including user needs, confidentiality issues etc.
- Strategies and policies

The first three bullets are very much related to IT developments. The last two bullets are mainly dissemination and subject matter related.

### 3. Conclusions and recommendations

It is emphasized that in order to get quality cubes the main focus should be on the work of the subject matter units and their business process. Thus, the project with cubes is not an IT project, rather the IT department should be seen as an enabler for the project to progress and succeed.

Based on the assessment it is recommended to:

#### **Initiate strategic measures**

The first step is to define a strategy and aims for the DWH. A strategic measure could be something like *“In 2020 all officially published data are derived from the DataWareHouse”*.

It should be decided when and with which data the DWH can be launched to the public, i.e. minimum requirements for the content and functionality of the DWH at the time launching.

#### **Develop or install tools**

It is necessary to develop or install tools to:

- Insert Code lists etc. (e.g. from the Dictionary system)
- Define cubes
- Load data (from existing data sources)
- View data (internally, before release and externally)

#### **Initiate a pilot project**

A pilot project should be initiated, involving subject matter units, dissemination and IT. A goal for the pilot has to be defined. This could be a number of specific statistics from selected subject areas that is to be created in the DWH.

#### **Create guidelines and policies**

The pilot project is supposed to help identifying needs for guidelines, policies and procedures. Guidelines should be created concerning the design of data cubes, naming conventions, rules for confidentiality etc. Furthermore, policies related to the release of data and updating procedures have to be put in place. A key element is to define the sharing of responsibility for the different parts of the process among the different parties involved (IT, Dissemination and subject matter units).

Investigate license costs and In-house development costs for the project.

### 4. Road map

The four items in the recommendations above is to be considered as a draft road map. Dates and persons responsible for each item should be added by the ICBS.

## **Annex E6.1.1 Terms of Reference**

### **Activity E6.1**

Integration with data warehouse

**20-23 October 2014**

#### **Mandatory results and benchmarks related to activity E6.1**

<b>Mandatory result</b>	<b>Benchmark</b>
<b>MR25.</b> Establishment of rules for updating of the website, including linkage of the data warehouse to the website	Rules and linkages to data warehouse established by 6 <sup>th</sup> project quarter

#### **Subject / purpose of activity E6.1**

Integration of ICBS website with data warehouse with particular focus on the development and application of updating rules and procedures regarding the data warehouse.

#### **Expected output from activity E6.1**

Mission report with recommendations regarding the integration between the ICBS website and the data warehouse.

## Annex E6.1.2 Agenda

Date	Time	Event
Mon 20/10	09:00	Welcome, introduction and objectives of the mission
	09:30	ICBS: Description of the actual status regarding the dissemination databases and other dissemination tools.
	10:30	Coffee break
	11:00	ICBS: <ul style="list-style-type: none"> <li>• Introduction to discussion of what ICBS' (unofficial) vision regarding the electronic dissemination is, as formulated in the preliminary dissemination strategy</li> <li>• The IT strategy as part of ICBS (preliminary) general strategy, and its connection to the dissemination strategy.</li> </ul>
	12:15	Lunch
	13:30	Statistics Denmark: <ul style="list-style-type: none"> <li>• Demonstration of Statistics Denmark's electronic dissemination</li> <li>• The general "single source" principle - discussion of the IT architecture behind Statistics Denmark's electronic dissemination</li> <li>• PC-Axis/PX-Web and Colectica: pre-defining cubes in Colectica and other integration</li> </ul>
	16:00	End of day 1
Tue 21/10	09:00	Summary from day 1
	09:15	Statistics Denmark ( <i>continuation from Monday</i> ):
	10:30	Coffee break
	11:00	Outlining possible alternatives for the development of an integrated system for ICBS's electronic dissemination: the IT architecture and accompanying updating procedures and rules
	12:15	Lunch
	13:30	<i>Cont. of discussion</i>
	16:00	End of day 2
Wed 22/10	09:00	Summary from day 2
	09:15	Discussion: revisiting ICBS's dissemination strategy in the light of the previous days' discussion
	10:30	Coffee break
	11:00	Discussion: Drafting a work-plan (road map) for the development of the IT architecture and accompanying updating rules and procedures
	12:00	Lunch
	13:15	<i>Cont. of discussion</i>
	16:00	End of day 3
Thu 23/10	09:00	Report writing and preparations for debriefing
	13:00	Debriefing
	15:00	End of meeting



### **Annex E6.1.3 Persons met**

## Annex E6.1.4 Mini-Pilot with ICBS data

Example with data „as is” from the Informix Times Series database. Two dimensions with mixed content and mixed units of measurement:

Students in schools by Time and Sex					
	Total	Boys	Girls	Hebrew	Arabic
1999	14,119.0	80.4	19.6	77.2	22.8
2000	13,493.0	81.7	18.3	77.4	22.6
2001	12,930.0	80.5	19.5	77.5	22.5
2002	15,049.0	80.3	19.7	77.3	22.7
2003	14,414.0	80.2	19.8	75.4	24.6
2004	14,454.0	81.2	18.8	76.0	24.0
2005	14,251.0	80.9	19.1	73.9	26.1
2006	14,200.0	81.4	18.6	70.9	29.1
2007	13,485.0	82.1	17.9	68.7	31.3
2008	13,065.0	83.5	16.5	66.2	33.8
2009	13,339.0	83.1	16.9	65.0	35.0

  

<b>Time *</b>	<b>Sex *</b>
<input checked="" type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input checked="" type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>
Total 11 Selected 0	Total 5 Selected 0
1999 2000 2001 2002 2003 2004	Total Boys Girls Hebrew Arabic
Search <input type="text"/> <input type="button" value="➤"/>	Search <input type="text"/> <input type="button" value="➤"/>
<input type="checkbox"/> Beginning of row	<input type="checkbox"/> Beginning of row

Example of the same data in a hypertetic “to-be” data cube. Five dimensions with possibility to cross variables such as “sex” and “language”. Same data are available but the example shows which data are missing (not available from time series data):

<b>persons or pct *</b>	<b>schools *</b>	<b>sex *</b>	<b>language *</b>	<b>Time *</b>
	<input checked="" type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input checked="" type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input checked="" type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input checked="" type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>
Total 2 Selected 0	Total 4 Selected 0	Total 3 Selected 0	Total 3 Selected 0	Total 11 Selected 0
Number of students Per cent	Total Schools under ministry Collegues Universities	Total Boys Girls	Total Hebrew Arabic	1999 2000 2001 2002 2003 2004
	Search <input type="text"/> <input type="button" value="➤"/>	Search <input type="text"/> <input type="button" value="➤"/>	Search <input type="text"/> <input type="button" value="➤"/>	Search <input type="text"/> <input type="button" value="➤"/>
	<input type="checkbox"/> Beginning of row	<input type="checkbox"/> Beginning of row	<input type="checkbox"/> Beginning of row	<input type="checkbox"/> Beginning of row

# Students in schools by schools, persons or pct, sex, language and Time

	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009
Schools under ministry											
Number of students											
Total											
Total	14,119.0	13,493.0	12,930.0	15,049.0	14,414.0	14,454.0	14,251.0	14,200.0	13,485.0	13,065.0	13,339.0
Hebrew	*	*	*	*	*	*	*	*	*	*	*
Arabic	*	*	*	*	*	*	*	*	*	*	*
Boys											
Total	*	*	*	*	*	*	*	*	*	*	*
Hebrew	*	*	*	*	*	*	*	*	*	*	*
Arabic	*	*	*	*	*	*	*	*	*	*	*
Girls											
Total	*	*	*	*	*	*	*	*	*	*	*
Hebrew	*	*	*	*	*	*	*	*	*	*	*
Arabic	*	*	*	*	*	*	*	*	*	*	*
Per cent											
Total											
Total	*	*	*	*	*	*	*	*	*	*	*
Hebrew	*	*	*	*	*	*	*	*	*	*	*
Arabic	*	*	*	*	*	*	*	*	*	*	*
Boys											
Total	80.4	81.7	80.5	80.3	80.2	81.2	80.9	81.4	82.1	83.5	83.1
Hebrew	*	*	*	*	*	*	*	*	*	*	*
Arabic	*	*	*	*	*	*	*	*	*	*	*
Girls											
Total	19.6	18.3	19.5	19.7	19.8	18.8	19.1	18.6	17.9	16.5	16.9
Hebrew	*	*	*	*	*	*	*	*	*	*	*
Arabic	*	*	*	*	*	*	*	*	*	*	*

## **Annex E6.1.5 Relevant links**

### **PC-Axis web site**

Documentation and scripts for the DWH model:

[http://www.scb.se/sv\\_/PC-Axis/Documentation/SQL-Metadatabase/](http://www.scb.se/sv_/PC-Axis/Documentation/SQL-Metadatabase/)

Download link for PX-Web software (for data viewing):

[http://www.scb.se/sv\\_/PC-Axis/Programs/PX-Web/PX-Web-2014/](http://www.scb.se/sv_/PC-Axis/Programs/PX-Web/PX-Web-2014/)

### **DST links**

Main web page and statbank (DWH):

<http://www.dst.dk>

<http://www.statbank.dk>

<http://m.statbank.dk> (mobile version – for mobile browsers only)

API for DWH:

<http://api.statbank.dk> (documentations)

<http://api.statbank.dk/console> (helping tools for generating requests)

### **Twinning links**

<http://www.dst.dk/israel>