



## TWINNING CONTRACT

**CRIS 2015/370-467**



# **Support to the Israeli Central Bureau of Statistics in Improving the Quality of Official Statistics**

## **MISSION REPORT**

**on**

**Component B**

**Micro Data services to researchers**

**Activity B.8**

**Confidentiality and Output Control**

*Implemented by*

**Mr. Eric Schulte Nordholt**, Adviser on Statistical Disclosure Control, Statistics Netherlands

**Jerusalem**

23-26 September 2018

Version: Final



**Authors' name, address, e-mail**

*Mr. Eric Schulte Nordholt*  
*Adviser on Statistical Disclosure Control*  
*Statistics Netherlands*  
*P.O. Box 24500*  
*2490 HA The Hague*  
*The Netherlands*  
[e.schultenordholt@cbs.nl](mailto:e.schultenordholt@cbs.nl)

**Table of contents**

Executive Summary .....	5
1. General comments.....	6
2. Assessment and results.....	7
3. Conclusions and recommendations .....	9

**Annexes produced for the mission (external to the report):**

Annex B8 – 1 Terms of Reference
Annex B8 – 2 Program
Annex B8 – 3 Persons met
Annex B8 – 4 Opening Statistical Disclosure Control
Annex B8 – 5 General introduction to Statistical Disclosure Control
Annex B8 – 6 Tables general
Annex B8 – 7 Tables methods
Annex B8 – 8 Introduction to Tau-ARGUS
Annex B8 – 9 Microdata protection general
Annex B8 – 10 Microdata protection methods
Annex B8 – 11 Introduction to Mu-ARGUS
Annex B8 – 12 Legal issues and access to microdata
Annex B8 – 13 Output checking

## List of Abbreviations

BC	Beneficiary country
EU	European Union
ICBS	Israeli Central Bureau of Statistics
IT	Information technology
MUC	Micro-data Under Contract
MS	Member State (of the EU)
NSS	National statistical system
OECD	Organisation for Economic Cooperation and Development
PCS	Public Council for Statistics
RAS	Remote Access System
PUF	Public Use File
RDC	Research Data Centre
SO	Statistics Ordinance
TA	Technical Assistance
ToR	Terms of Reference
UNECE	United Nations Economic Commission for Europe

## Executive Summary

*As part of the EU Twinning Project IL/12 CRIS 2015/370-467 Eric Schulte Nordholt (Statistics Netherlands) visited the Israeli Central Bureau of Statistics (ICBS) in the period 23-26 April 2018. It concerns Component B (Micro Data services to researchers) and Activity B.8 (Confidentiality and Output Control).*

*Main conclusions and highlights from findings:*

- *ICBS is in the legal position to centralise the access facilities for researchers to get an efficient way to give researchers access to secure use files;*
- *The amount of work to be done to set-up and maintain these centralised facilities should not be underestimated;*
- *Training of ICBS staff members will be an essential part of the activities to make the centralised facilities a success.*

## 1. General comments

This mission report was prepared as part of the Twinning Project “*Support to the Israeli Central Bureau of Statistics (ICBS) in Enhancing the Quality of Official Statistics*”. This was the seventh Twinning activity of Component B (Micro Data services to researchers) and focused to Confidentiality and Output Control.

The main activities of the mission were:

- Presentations by the Charlotte Nielsen (Resident Twinning Adviser of Statistics Denmark) and ICBS on Monday 23 April 2018 to get acquainted with the situation at the ICBS;
- Presentations on Statistical Disclosure Control and Output Checking by Eric Schulte Nordholt on Monday 23 April, Tuesday 24 April and Wednesday 25 April followed by discussions;
- Discussions of Eric Schulte Nordholt with different ICBS colleagues about Micro Data services to researchers on Wednesday 25 April and Thursday 26 April.

The mission assisted the ICBS and the Twinning Project experts in getting input for Indicator IB7: Guidelines for output approved by the Confidentiality Committee to be completed by 9th project quarter (end of May 2018).

I would like to express my thanks to all officials and individuals met for their kind support and for the valuable information they provided, which highly facilitated the experts' work. I want to thank especially Charlotte Nielsen who was as Resident Twinning Adviser a very good host of this mission and helped me with lots of practical aspects.

The views and observations stated in this report are those of the consultants and do not necessarily represent the views of Statistics Netherlands, ICBS or Statistics Denmark.

## 2. Assessment and results

Starting point of the mission was the list of activities mentioned in the Terms of Reference (ToR). All activities mentioned in the ToR have taken place on 23-26 April 2018. During those days there was no time to write the report and that has thus been done after the mission.

A description of the current situation regarding researcher's access to microdata and output control has been described in the ToR and was presented on Monday 23 April. After this presentation the current situation at ICBS was also discussed in the beginning of the mission. At the end of the mission the situation was assessed. Then some conclusions were drawn and recommendations were made that can be found in section 3 below.

During the mission ten presentations on Statistical Disclosure Control and Output Checking were given. The knowledge at ICBS of Statistical Disclosure Control in general and in Output Checking in particular has been improved during the mission. All presentations were attended by many staff members of the ICBS. Many questions were asked and answered. After each presentation the topic was discussed more generally.

On Wednesday 25 April the vision and recommendations could not be presented because of the absence of a number of ICBS managers. The final discussions on Thursday 26 April made clear that a lot of work on the access facilities has to be done at ICBS, but that it is worth taking those efforts to improve the access to microdata in a secure way.

It was advised to follow the rules for output checking according to the rule of thumb model as defined in the Data without Boundaries project (see e.g. [https://ec.europa.eu/eurostat/cros/system/files/dwb\\_standalone-document\\_output-checking-guidelines.pdf](https://ec.europa.eu/eurostat/cros/system/files/dwb_standalone-document_output-checking-guidelines.pdf)).

The rule of thumb model is based on clear and simple (strict) rules. Because these rules differ only slightly for different classes of output, an overall rule of thumb can be established.

This overall rule of thumb is presented first. Thereafter lists of unsafe and safe output according to this rule are given.

The overall rule of thumb has four parts:

1. **10 units:** all tabular and similar output should have at least 10 units (unweighted) underlying any cell or data point presented. A common term for such a rule is a threshold rule (the cell count must exceed a specified threshold).
2. **10 degrees of freedom:** all modelled output should have at least 10 degrees of freedom and at least 10 units should have been used to produce the model.  
Degrees of freedom = number of observations - number of parameters - other restrictions of the model
3. **Group disclosure:** in all tabular and similar output no cell can contain more than 90 % of the total number of units in its row or column to prevent group disclosure. Group disclosure is the situation where some variables in a table (usually spanning variables) define a group of units and other variables in the table divulge information that is valid for each member of the group. Even though no individual unit can be recognized, confidentiality is breached because the information is valid for each member of the group and the group as such is recognizable.
4. **Dominance:** in all tabular and similar data the largest contributor of a cell cannot exceed 50 % of the cell total.

Note that the values of the parameters mentioned above (e.g. the threshold of 10 units) are a convention with the purpose to be able to quickly filter out safe output. Output not passing these rules, is not necessarily disclosive, but needs further attention.

All output can be classified into a limited number of output types. Then each type of output can be marked as either generally safe or generally unsafe. The following lists result.

Unsafe:

- Frequency tables
- Magnitude tables
- Maxima, minima and percentiles (incl. median)
- Means, indices, ratios, indicators
- Graphs: pictorial representations of actual data
- Estimation residuals

Safe:

- Mode
- Concentration ratios
- Higher moments of distributions (incl. variance, covariance, kurtosis, skewness)
- Linear regression coefficients
- Non-linear regression coefficients
- Summary and test statistics from estimates ( $R^2$ ,  $\chi^2$ , etc.)
- Correlation coefficients
- Factor analysis
- Correspondence analysis

If data are marked as safe they can be released according to the overall rule of thumb. If data are marked as unsafe they cannot be released or a thorough check has to be executed before release is possible (this is called the principle based approach).



### 3. Conclusions and recommendations

The following conclusions related to the activities mentioned in the ToR can be listed:

- ICBS is in the legal position to centralise the access facilities for researchers to get an efficient way to give access to secure use files;
- The amount of work to be done to set-up and maintain this facility should not be underestimated;
- A basic level of knowledge of Statistical Disclosure Control and Output Checking is available in the ICBS on which can be relied;
- To make the new centralised facility a success a number of actions have to be taken as described in the recommendations below.

The following recommendations for moving forward are made:

- Improve the level of Statistical Disclosure Control and Output Checking by joining existing ESTP courses (small number of trainees) and courses at ICBS (larger number of trainees);
- Make the access facilities for microdata a joint responsibility of ICBS and the researchers that make use of the facilities;
- Make output checkers aware that the checking is context specific;
- Check whether the four eyes principle (a SDC expert and a context expert check the difficult output independently) can be followed;
- Invest in the consistency of the output checks by discussions and organising seminars;
- Be realistic in the time it costs to check the output;
- Try to avoid large numbers of outputs offered to check by putting a price on checking the output;
- Try to diminish the number of outputs further by offering researchers to write their complete reports at the facility and check only their final output;
- Give the rules for output checking a formal status and publish them on the website as part of the information how researchers can get access to microdata (see for an example in another country <https://www.cbs.nl/en-gb/our-services/customised-services-microdata/microdata-conducting-your-own-research>).