



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.3

Organizational structure, procedures and processes

Implemented by

- Mr. Ivan Thaulow, MS Component Leaders Head of Division – Research Services, Statistics Denmark, ith@dst.dk
- Mr. Jørn Korsbø Petersen Chief Advisor Research Services, Statistics Denmark, jkp@dst.dk

Jerusalem

September 26-29 2016

Version: Final

Authors' names, addresses, e-mails

*Mr. Ivan Thaulow,
MS Component Leader, Head of Division – Research Services,
Statistics Denmark
Sejrøgade 11
DK-2100 Copenhagen Ø
Denmark
Tel: +45 39 17 31 31
Email: ith@dst.dk*

*Mr. Jørn Korsbø Petersen
Chief Advisor Research Services
Statistics Denmark
Sejrøgade 11
DK-2100 Copenhagen Ø
Denmark
Tel: +45 39 17 32 33
Email: jkp@dst.dk*

Table of Contents

Executive Summary	5
1. General comments.....	6
2. Assessment and results.....	7
3. Conclusions and recommendations	13

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

Annex B3 - 1 Terms of Reference
Annex B3 - 2 Meeting program for B3
Annex B3 - 3 Persons met
Annex B3 - 4 Overview of Micro Data Access with Focus on European Countries (BC Presentation)
Annex B3 - 5 Step by Step Procedures for researcher's access to Microdata from DST (Background paper)
Annex B3 - 6 Process 1 - Recognition as a Research Entity (BC Presentation)
Annex B3 - 7 Process 2 - Approvals of Researchers (BC Presentation)
Annex B3 - 8 Process 3 - Project Approval (BC Presentation)
Annex B3 - 9 Process 4 - The Data Delivery Process (BC Presentation)
Annex B3 - 10 Process 5 - Output control (BC Presentation)
Annex B3 - 11 Guidelines for Transferring Aggregated Results (Background paper)

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

This mission report deals with the third mission within the Twinning Project "Support to the Israeli Central Bureau of Statistics (ICBS) in Enhancing the Quality of Official Statistics". The mission was devoted to ***Organizational structure, procedures and processes*** within Component B: Micro Data Services to Researchers of the Project.

The Mission was initiated with ICBS presented current status and the visions of the future research service as well as progress since last mission. In brief, the vision is to establish a dedicated research service unit that will be responsible for coordination and tracking all processes involved in research projects as well to provide more simple and lean procedures and processes both for the researchers as well as internally within ICBC. Furthermore, the vision is to provide a remote access solution (virtual access facility) with a design that covers existing solutions with research rooms files (RR)¹ as well a micro-data under contract files (MUC)².

Since the last mission effort has been dedicated to design technical requirement for a remote access system including budget as well as designing an administrative system enabling ICBS to track status of all research projects. In addition work on obtaining better estimates for the cost associated to research projects has been going on. In parallel the potential for financial support for both a future remote access system as well as for services are examined.

Based on a number of presentations and meetings between relevant ICBS staff and the Danish Twinning experts, solutions for organizational structure, procedures and processes was assessed and discussed. Obstacles for the vision were discussed such as case to case approval of data transfer between subject units and case by case approval of the confidentiality committee – procedures that today prevent the forming of a centralized research service at ICBS.

Furthermore, it was decided that in the transition process, regular monthly meetings between (a) the Confidentiality Committee, (b) selected staff from the subject units and (c) the International Relations and Statistical Coordination Department will be organized. Each meeting should have a central subject on the agenda. The first meeting will concentrate on output control. The MS Experts also came up with a number of conclusions and recommendations of which the most central were the following:

- To create a mechanism that transfers economic resources to the subject units when they carry out services to researchers
- To transfer resources to ensure sufficient resources for participating actively in the work group and steering committee creating the centralized research service unit
- That the Confidentiality Committee become an advisory board, creating policy and guidelines for access to micro-data rather than being involved all research projects
- To remove internal barriers not bound by law for data transfer between subject units
- To build a standardized SAS-library containing the most used micro-data in research projects
- To create internal as well as external guidelines for output control

¹ Research rooms files or 'secure-use files' means confidential data for scientific purposes to which no further or only limited methods of statistical disclosure control have been applied

² Micro-data under Contract (MUC) or scientific-use files' means confidential data for scientific purposes to which methods of statistical disclosure control have been applied to reduce to an appropriate level and in accordance with current best practice the risk of identification of the statistical unit;

1. General Comments

This mission report was prepared as part of the the Twinning Project "Support to the Israeli Central Bureau of Statistics (ICBS) in Enhancing the Quality of Official Statistics". It was the third mission and it was devoted to Organizational structure, procedures and processes for research services within Component B: Micro Data services to researchers of the project.

The purposes of the mission were based on a number of presentations on the central processes in relation research projects to discuss organizational structure and procedures.

Processes and services in focus were:

- Approvals
 - Recognition as a research entity (authorization process)
 - Approval of researchers
- Project approvals
- The data delivery process
- Output control
- Pricing and contract

The presentations assisted the ICBS and the Twinning Project experts in forming the basis for developing a long-term plan for allowing researchers remote access to de-identified micro-data in Israel.

The experts 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 Israel and which highly facilitated the work of the experts.

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

2. Assessment and Results

During the mission the following activities took place; cf. *Terms of Reference (Annex B3- 1)*:

Since the last mission effort has been dedicated to design technical requirement for a remote access system including budget as well as designing an administrative system enabling ICBS to track status of all research projects. In addition, work on obtaining better estimates for the cost associated to research projects has been going on. In parallel the potential for financial support for both a future remote access system as well as services are examined.

2.1 The Visions for a future Research Service Unit at ICBS

ICBS presented current status and the visions of the future research service. The vision for a future research service at ICBS can be summarized as:

- A centralized research service unit that will handle all processes including administrative processes such as approval of researchers, counseling and evaluation of project descriptions according to the “need to know” principle, extraction of data from the SAS-library and giving researchers access to data such as MUC-files or micro-data in research rooms – later by remote access.
- A central research service unit has access to both metadata and micro-data from ICBS that is relevant for research purpose.
- Data are stored in a harmonized library securing an efficient service of the researcher’s requests.
- To provide a remote access solution (virtual access facility) with a design that covers existing solutions with research rooms files (RR)³ as well a micro-data under contract files (MUC)⁴.

Obstacles for the vision were discussed such as case to case approval of data transfer between subject units and case by case approval of the confidentiality committee – procedures that today prevent the forming of a centralized research service at ICBS.

2.2 Overview of Microdata access for researchers provided in European countries

In order to set the scene for the missions and in order to get a common understanding of terms and definitions the Resident Twinning Advisor gave a presentation on micro-data access for researchers provided in European countries.

Processes in common were highlighted but not at least some of the processes that are handled differently either due to differences in either national legislation or maturity of the organization was pointed out (Annex B3 – 4).

2.3 Meeting with the Confidentiality Committee

Initially the vision for the future research service was presented to the Confidentiality Committee by Ms. Sigalit Mazeh, BC Component Leader, Director, International Relations and Statistical Coordination Department. Their active participation in the planning of the future research service was discussed.

³ Research rooms files or ‘secure-use files’ means confidential data for scientific purposes to which no further or only limited methods of statistical disclosure control have been applied

⁴ Micro-data under Contract (MUC) or scientific-use files’ means confidential data for scientific purposes to which methods of statistical disclosure control have been applied to reduce to an appropriate level and in accordance with current best practice the risk of identification of the statistical unit;

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

This was followed up by Mr. Ahmad Hliheli, Head of the Confidentiality Committee, who on behalf of the committee presented the four major challenges from their point of view, in brief:

- The lack of guidelines for the “need to know”-principle
- The lack of guidelines for checking researchers output
- The use of MUC-files contra RR-files
- Use of synthetic data

The need for guidelines was discussed as well as the relevance of continuing producing PUF-files, when the researchers have remote access to MUC-files and RR files from numeration access locations.

2.4 Meeting with the Head Accountant Ms. Orit Levin

Calculation of the total and accurate costs for ICBS for giving researchers access to microdata is under progress. Calculation of the total costs is important to ensure correct prices for the services provided to researchers. Presently it is known that ICBS does not charge the researchers for the actual cost (time used by staff, rent for locations, IT-tools and programs). Preliminary estimates indicate that actual costs are at least double.

2.5 Meeting with the subject unit staff

The visions for the future research service were presented to selected subject unit staff since their active participation and engagement in the planning of a more centralized service to researchers is essential. A very engaged discussion took place. The major challenges experienced by the subject units point can be summarized as:

- *Lack of Guidelines from the Confidentiality Committee*
 - Lack of Guidelines for how to interpret the need to know principle
 - Lack of Guidelines for output control
 - Lack of Guidelines for processes (standardized processes).
- *Lack of service agreements - internally as well as externally*
 - What can be expected?

- *Improved communication*

The subject units demanded a better dialog with (a) the International Relations and Statistical Coordination Department, (b) Confidentiality Committee and (c) the Committee of Data Transfer.

They pointed out that dialogue is important in particular in relation to the current situation and in relation to the development of a new structure for the research service and not least for the transition period for establishing a centralized research service unit.

- *Estimation of time needed for each project (pricing – calculated in full days).*
 - Estimation of time is particularly difficult when it comes to new researchers that the subject units have never worked with before, especially estimation of time needed (expected) for consulting with researchers as well as the Confidentiality Committee is a challenge. Also time estimation of output control is seen as a major challenge - no guidelines or agreements of number and size of files for output control is in place. It was suggested to output control should be 'priced' separately
- *Time constrains*
 - Providing service to the researchers is often something that comes on top of other obligations from the subject unit such as e.g. publications that are regarded as the main activity of the subject unit so service to researchers will often end up being extra

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

working hours for the employees. Also performance payment is based on timeliness of publications rather than providing service to the researchers.

- Citation 1 from Adnan Mansur from Department of Education: “The budget has already been concluded in the beginning of the year – services to the researchers is extra work not considered in the budget – with the growing demand from the researchers I get more work. I need more people!!”
- Citation 2 from Adnan Mansur: “A request for research data might suddenly come in right before a publication were we are really busy”
- *Lack of catalog of data available for researchers*
 - Citation from Adnan Mansur from Department of Education: “In order to integrate his data on education with data from other internal ICBS sources he need to learn about the other data which is time consuming with a lot of unknown factors. No Metadata available
- *Transfer of data between subject units is challenging*
 - Transfer of data between subject units within ICBS always requires approval from the Committee of Data Transfer. The process is time consuming and lack coordination
 - Citation from Adnan Mansur from Department of Education: “The administrative burden of combining data from several subject units is too big”
 - At the meetings it was mentioned that good guidelines and procedures are in place when it comes to access to external data, but this is entirely lacking when internal ICBS data has to be combined.
 - Citation from Yafit Alfandari, Consumption and Finance Unit from the Department of Micro-Economic Statistics “The process of combining data from just two units can take up to three months”

Understanding the researchers

- The requests from researchers are complex and complicated; The research proposals received are not always described in enough detailed–clarification from the researchers is needed before the subject unit can proceed to extract data and construct metadata
- It is a challenge to guide researchers using applications other than SAS for example. R, STATA etc

Besides pointing out challenges, the meeting also functioned as a forum for knowledge sharing, thus Mark Feldman, Director of the Labour Force Unit, informed that their unit has created a research dataset they use as the basis for providing data to research projects. This means that they can deliver data within a few weeks after receiving the request from the researchers (compared to up to several months for some units) provided that researchers only want data from LSF. This fact was supported by Anat Katz-Avram from the International Relations and Statistical Coordination Department who is currently responsible for coordinating and following up on all research services processes. This was followed by discussing the need for standardized files MUC.

It was concluded that in the transition process regular monthly meetings between (a) the Confidentiality Committee, (b) selected staff from the subject units and (c) the International Relations and Statistical Coordination Department will be organized. Each meeting should have a central subject on the agenda. The first meeting will concentrate on output control.

2.5 *MS presentations*

The presentation was built around the five central processes in focus during the mission and that are listed below:

- Approvals
 - Recognition as a research entity (authorization process)
 - Approval of researchers
- Project approvals
- The data delivery process
- Output control
- Pricing and contract

Authorization of research environments at Statistics Denmark

The process of authorization of research environments at Statistics Denmark was presented and discussed (Annex B3 – 6)

At statistics Denmark access to micro data can be granted to researchers and analysts from research environments pre-approved by Statistics Denmark. Authorizations can be granted to public research and analysis environments (e.g. universities, sector research institutes, ministries, etc.), to research organizations as a part of a non-profit foundation as well as private companies under certain conditions. Statistics Denmark will not grant authorization to single persons.

In order get an authorization an application needs to be submitted to the Research Service Unit at Statistics Denmark. The application should include 1) a description of the research environment; 2) name and contact information for a responsible manager of the authorization (often the head of the department).

Statistics Denmark will evaluate the application carefully and make a concrete assessment of the applicant's reliability as a data recipient (e.g.: ownership, educational standard among the staff and the research experience). If the institution is approved for access to micro data an authorization contract between Statistics Denmark and the research institution/environment is signed by the National Chief Statistician and the responsible research manager at the institution (For details please see appendix A).

The most important point in the authorization contract is that the person signing the authorization agreement on behalf of the research/analysis environment undertakes the role of personally supervising that the confidentiality of data is always kept. The researcher may not, directly or indirectly, download the dataset or any datasets derived there from. Furthermore, individual records may not be printed, and all outputs must be aggregated to an extent that eliminates any risk of direct or indirect identification of persons or enterprises. The researcher may not attempt to make such identification.

Approval of researchers

The process of approving researchers at Statistics Denmark was presented and discussed (Annex B3 – 7)

Researchers have to sign an agreement with Statistics Denmark. The agreement states that all work on micro data must take place on the research server and no attempts to identify people or enterprises – or to remove micro data must be made. *Approval of project*

The process of approving a project at Statistics Denmark was presented and discussed (Annex B3 – 8)

Access to data is given according to the “need to know”-principle, according to which the researcher can only be granted access to data needed to fulfil the research purpose.

The proposal needs to cover the following paragraphs:

- Description of purpose of the research project
- Population to be studied (need to know)
- DST registers included in the study (need to know)
- External data (survey data or data from external registers)
- Register year
- Who will have access to data

In order to assist the researchers in selecting the registers of relevance, comprehensive descriptions of most of the registers provided by Statistics Denmark. The declarations describe quality and content of data as well as information about purpose, history, collection method, accuracy, reference period, and comparability over time.

The three processes above at Statistics Denmark supported by a newly implemented administrative system that ensures a consistent registration and administration of the interrelated processes: authorization, approval of researchers and approval of projects. Behind this system lies a model illustrated in figure 1:

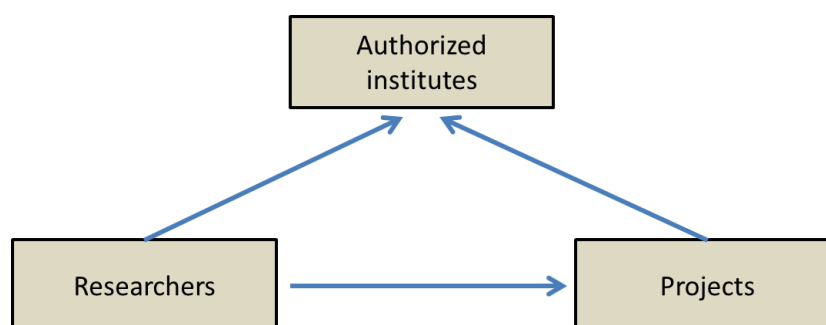


Figure 1: Illustration of Statistics Denmark's model for administration of research institutes, projects and researchers

The model ensures that researchers can only have access to projects owned by an authorized institution and that a researcher must be affiliated with an authorized institution before being granted access to data from a specific project. If one of these terms is not met then access cannot be granted. E.g. a new research project can only be created if the project is owned by an authorized institute and a researcher must have signed an agreement with an authorized institution, before the researcher can have access to data from the research project.

The data delivery process

The largest process of preparing data for research projects at Statistics Denmark was presented and discussed (Annex B3 – 9)

This process also contain sub-processes, calculation of price and contract writing was presented and discussed.

Based on the project description the Research Service Unit will prepare a detailed data extraction protocol, including detailed lists of variables as well as sets of values, if they are of relevance. Descriptions of all variables can be found in our documentation system TIMES, which is also published on the internet, so far in Danish only.

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

The 600 mostly used variables have very detailed variable descriptions, including data breaches as well as appendices with special interest.

Based on the data extraction protocol a price will be calculated. Ten hours of free initial advisory services are given for new projects managed by public research and analysis environment. Additional charge for disk storage in the case of large-scale projects and log-on charge for private companies will be charged. Finally, a contract stating price and expected time of delivery is sent.

Data extraction from a central SAS library managed by the Research Service Unit will be initiated shortly after Statistic Denmark has received a signed contract. Data are anonymized by means of a project-specific key before data are entered into the researcher computer (all identifying variables such as CPR numbers, addresses etc. are replaced by project specific random num-bers). When research data for a project have been prepared by the Research Service Unit, the data are transferred to the research server where remote access is given via the Internet. Communications via the Internet is protected and encrypted, which secures against unauthorized access.

Output control

The process of output control including sanctions by data breach was presented and discussed (Annex B3 – 10).

Researchers can send home aggregated output from the research computers to their own mail. Output is not checked prior to export, but all output send home by the researcher is registered by Statistics Denmark. A random sample of the exported output is checked by Research Service. In the event of a data breach is the institution sanctioned.

Research Service is currently developing software that automatically will check the output for microdata – primarily de identified key variables – before output is send home. If microdata are found in the output the researcher receives a warning.

The Danish Guidelines for sending home aggregated output from the research computers can be found in Annex B3 – 11.

3. Conclusions and Recommendations

The actions planned for this activity were carried out according to the programme in the ToR.

The recommendations in this activity assume that the recommendations in activity 1 and activity 2 are met. We find it e.g. important that the recommendation of the establishment of Working Group and a Steering Committee and the recommendation to make an organizational reorganization with one dedicated unit, the Research Service Unit, that will be responsible for coordination and tracking all processes involved in research projects are met, cfr. 3.1.1 and 3.1.2 in Mission Report, Activity B.1.

Since the recommendations in activity 1 and active 2 focus primarily on the overall organization of a future research service the recommendations from this activity will focus more on the processes carried out in a centralized research service as presented by MS.

3.1. General recommendations

Before the transfer of research services from the subject units to the centralized research unit we recommend creation of a mechanism that transfers economic resources to subject units with a present high research workload. We also recommend transfer of resources to ensure sufficient resources for participating actively in the work group and steering committee creating the centralized unit.

3.2. Processes of approval

It is the impression of MS that the process of approval of researchers is well functioning at ICBS. However we recommend, that ICBS consider if it is beneficial to implement the Danish administrative model as described in paragraph 2 with responsible authorized institutions e.g. for a consistent handling of data breaches.

3.2.1 The confidentiality committee becomes an advisory committee

The Confidentiality committee approves today all the projects before researchers can have access to data. As a part of recommendation 3.1.3, Leaning of Workflows in Mission Report B1, the MS experts recommend that the confidentiality committee (CC) becomes an advisory committee, creating policy and guidelines for access to microdata, and that the approval of the specific projects is transferred to the Research Service Unit. Issues of general importance must be addressed to CC by the Research Service Unit.

3.3. More efficient data preparation process

It is of central importance for an efficient case by case project handling that as many processes as possible is controlled by the Research Service Unit only taking into account legislative obstacles. We therefore recommend that internal processes not bound by law that prevents an efficient data preparation process is reconsidered and if possible removed, cfr. 3.1.7 Legal Framework in Mission report B1.

To support a more efficient data preparation process for a centralized Research Service Unit we recommend:

- Removal of internal procedures not bound by law that creates barriers for data transfer between subject units – e.g. by help of the working group including experienced people from the subject units
 - Recommend guidelines for clear and transparent rules for internal and external handling of microdata

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

We recommend that all relevant data relevant for research in ICBS should be made accessible for the Research Service Unit. Based on experiences from Statistics Denmark this would ensure an efficient service of the researchers, build up a comprehensive knowledge of data among the staff in the Research Service Unit, and improve programming skills. It will be difficult for a centralized Research Service Unit to function efficient, if data are not centralized.

- Recommend buildup of a standardized SAS-library with the microdata needed for research purposes from the subject units.
 - As a first step mapping of microdata relevant for research to include in a SAS-library across subject unit – and codebooks for documentation
 - Develop procedures for standardization for efficient creation of research-data
 - Same variable names for ID-variables across datasets
 - Same use of SAS-formats across datasets
 - Develop procedures for updating the SAS-library with data from the subject units
- Recommend development of a policy for adding new relevant data to the SAS-library
- Recommend development of documentation of SAS-library for internal and external use
- Recommend internal guidelines for handling data for researchers describing standardized procedures and service level expected by the researchers

3.4.Output control

Today ICBS check all output the researchers want to bring home from the research laboratories. At the present level of security we recommend that this practice is continued. However in the future when experience is gained, possibly combined with remote access with an expected increase in output, we recommend that ICBS consider checking samples.

Presently as well as in the future we recommend internal as well as external guidelines for output control including as a minimum for the following subjects:

- What output must the researcher bring home
- What must the researcher not bring home
- How must the output be checked by the staff in ICBS
- Internal guidelines for a data breach
 - Include a clear definition of a breach for the researcher
- Sanctions for the researcher / research environment by a breach

To proactively preventing security breaches we recommend that ICBC develop a security course for researchers.

We also recommend studying the security systems in Holland and Finland