TWINNING CONTRACT

CRIS 2015/370-467

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

MISSION REPORT

on

Component B
Micro Data services to researchers

Activity B.1
Position analysis and presentation of experiences of micro data access for scientific purposes in Europe

Implemented by

- Mr. Ivan Thaulow, MS Component Leaders Head of Division – Research Services, Statistics Denmark, ith@dst.dk
- Mr. Bo Guldager Clausen Chief Advisor Division of IT, Statistics Denmark, bgc@dst.dk

Jerusalem

March 14-17 2016

Version: Draft no 6

הלשכה המרכזית לסטטיסטיקה
Central Bureau of Statistics
دائرة الإحصاء المركزية

STATISTICS DENMARK

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. Bo Guldager Clausen
Chief Advisor Division of IT
Statistics Denmark
Sejrøgade 11
DK-2100 Copenhagen Ø
Denmark
Tel: +45 39 17 38 15
Email: bgc@dst.dk

Table of Contents

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

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 first 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 position analysis and presentation of experiences of micro-data access for scientific purposes in Europe within Component B: Micro Data Services to Researchers of the Project.

Based on a number of presentations and meetings with the relevant staff of ICBS the current situation on research services at ICBS – as well as in Denmark and in other relevant countries - was described, discussed and assessed. From this the MS Experts came to a number of conclusions and recommendations of which the most central concerning **organization, workflow, business plan etc**. are the following:

The MS Experts recommend establishing a Steering Committee as well as a small inter-institutional working group that may ensure progress in the component between missions. The working group should include representatives of significant partners. The Steering Committee should preferably be established at a high level meeting of directors from the ICBS.

In ICBS there seems to be many internal players involved in each research project; however leadership and coordination seems to be missing. It is recommended that ICBS make an organizational reorganization with one dedicated unit that will be responsible for coordination and tracking all processes involved in research projects.

The division of labour and the major workflows concerning research service has to be agreed upon within the ICBS. Some confusion seems to be present. Far more transparency is needed about Research Service in ICBS.

Several work processes/workflows need to be more simple and lean. As many work processes/workflows as possible should be made automatic from the beginning.

At present, little is known about how much time ICBS actually spends on providing services to researchers - and where in the organisation the time is spent. The experts recommend implementing a documenting system in all ICBS for assessing the actual time used on research services as soon as possible.

Furthermore, the experts recommend that total and accurate/qualified calculations on costs, as well as income relating to research services, are made in order to make ICBS able to decide on the yearly economic result of the overall Research Service activities. Thus, a gathered business plan could be drafted.

Concerning the future IT solution, the experts especially recommend that the decisions on the server virtualization model (e.g. VMWare VDI, MS RDP) and the statistical software which shall be offered to researchers are made.

It is also recommended to make a draft of the IT design of the remote access solution and a draft list of requirements for automated output control too.

Finally, it is recommended that ICBS consider creating a service-desk for researchers.

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 first mission to be devoted to Position analysis and presentation of experiences of micro-data access for scientific purposes in Europe within Component B: Micro Data services to researchers of the project.

The purposes of the mission were to:

- Assess the current situation in Israel
- Outline initial recommendations to a program for providing researchers with access to micro-data in Israel (Organization, workflow, legal framework, IT Solutions)
- Prepare a detailed overview and time schedule (including deadlines) of the remaining activities in the component prepared (Taking the outcome of the proposal for possibly funding to establish a Remote Access Facility for MUC files into considerations)

The position analysis assisted the ICBS and the Twinning Project experts in getting an overview of the present situation regarding services to researchers and research environments in Israel, including administrative, technical, legal and technological aspects.

MS experts presented the background and the strategy behind how micro-data is provided in Europe with focus on solutions in Denmark and other Nordic Countries as well as in the EUROSTAT. Essential topics included the legal framework, administrative framework, process flow, business plan, technology, human resources, training and IT.

The position analysis thus established the current state of play regarding research services and micro-data in Israel. Furthermore, the position analysis contrasted the current situation with European best practices and thereby formed 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 B1- 2)*:

- ICBS presented the current status of researcher access to microdata. The presentations included:
    - *Microdata services to researchers at ICBS - Current Status, Goals and Objectives (Annex B1 - 4)*
    - *Access to microdata for research at ICBS - Legislation, current practice and initiatives for change (Annex B1 - 5)*
    - *Access to microdata – Current status on and future plans for IT (Annex B1 - 6)*

- MS presented the current status and future plans for micro-data access in Europe with focus on solutions in Denmark and other Nordic Countries as well as in EUROSTAT. The presentations included:

    - *General introduction to research services in Denmark including organizational structure (numbers, skills and competences) (Annex B1 -7)*
    - *Presentation of Technological infrastructures and security in Denmark (Annex B1 - 8)*
    - *Presentation of research services and micro data in Nordic countries and EUROSTAT (Annex B1 - 9A and Annex B1 - 9B)*
    - *Presentation of research services and micro data strategy in Denmark (Annex B1 - 10)*
    - *Workflows and processes including interactions with internal actors (subject units, IT, finance and administrative units) and external actors (Researchers, Ministries, data providers) (Annex B1 - 11A and Annex B1 - 11B)*
    - *Legal framework / necessary legal considerations for providing access to de-identified micro data in Denmark, authorization and approval of research projects (Annex B1 - 12)*
    - *Presentation of the business plan at Research services at Statistics Denmark - including Cost and pricing structure (Annex B1 - 13)*

Based on these presentations discussions on mission, vision and objectives with a focus on organization, workflow, legal framework, IT Solutions and business plans took place.

2.1 Current situation at ICBS assessed

ICBS had prepared comprehensive presentations giving the experts an overview of the current situation for researchers' access to micro-data at ICBS. In parallel some of the challenges that ICBS and the researcher find most significant were highlighted (Please find uploaded presentations at www.dst.dk/israel).

2.1.1 Status, Organization, Workflow, Business Plan and Legal Framework

Presently ICBS provide the following three types of micro-data files for researchers:

- ***PUF(Public Use Files)*** – defined as microdata that has been subject to statistical disclosure control methods that render the data non-confidential according to the legal and methodological standards applicable in Israel. Available to public under restrictive license. Data can be stored on any platform.
- ***'Scientific-use' MUC files (Microdata Under Contract)*** – defined as confidential microdata for research purposes that has been subject to statistical disclosure control methods, that together with an administrative, legal and logistical framework minimize risk of identification to a level allowing access to it by approved researchers from accredited research institutions, according to the legal standards applicable in Israel. Data could be accessed by recognized research Institutions via remote access although only available until 2015. Currently, these files are only available for use in one of the ICBS research rooms.
- ***'Research room' files*** – defined as anonymized confidential data for research purposes to which only limited or no further methods of statistical disclosure control have been applied. Data can only be accessed by approved researchers from accredited research institutions, from secure research rooms hosted by ICBS. All researchers are subject to legal and administrative restrictions, including the making of each researcher a special sworn employee of the ICBS.

Currently 55 active research projects are hosted by ICBS in its research rooms (37 in Jerusalem, 8 in Haifa, and 10 in the Bank of Israel remote access research room. In addition a research room in Tel Aviv is scheduled to open in later this April 2016. However, the number of research project there is unknown at present) In addition there are about 40 additional research projects in different stages of progress. In total about 100 researchers have access to the current 55 projects distributed among 25 different recognized research institutions (Academic institutions, other recognized research institutions, Governmental Institutions as well as Industry).

MUC and research room datasets are custom made for a fee according to *the need to know principle* and often integrate survey and administrative data. Data from all sources are combined and delivered to the researcher as just one dataset. The use of administrative data is subject to approval by the data provider. It is the responsibility of the subject matter department to prepare the dataset and vet the output (All outputs from MUC and research room files are subjected to output control).

Presently standard PUF and MUC files cost 100 NIS. For custom made data the cost is calculated on the basis of expected number of work days. The cost for research datasets range between 1,500 - 60,000 NIS per project depending on expected time needed to construct the dataset. The cost of using a research room is 3,100 NIS a year per project at present but ICBS plans to raise the amount. For the Bank of Israel remote access research room there is no annual charge, because that system uses a server hosted by ICBS but paid for by the Bank of Israel. It was stated that the current fee paid by the researcher does not cover the cost of providing micro-data to the researcher and research services, despite the fact that the exact cost of each project is presently unknown since there is no registration of time used by the involved partner.

The organization and workflow both internal within ICBS as well as for the researcher is rather complex because it involves several departments, units, committees and managers from ICBS and includes several steps of approval. Therefore it may take the researcher up to a year from the initial request for data access until the data is available.

As part of the outcome from this Mission, an overview of responsibility of each department and unit was prepared *(Annex B1 - 14)* and a workflow in Hebrew was presented and translated to English after the Mission *(Annex B1 - 15)*

The most relevant legislation includes The Statistics Ordinance and The Protection of Privacy Law, 1981, which were both carefully presented by the Legal Adviser of ICBS.

2.1.2 IT Solutions

At present ICBS provides microdata to researchers using three different IT solutions depending on the type of data.

Research rooms – for 'research room files' and currently also for 'MUC' files

ICBS have research rooms located in Jerusalem, Tel Aviv, Haifa and the Bank of Israel.

In Jerusalem there are 10 standalone Windows 7 computers placed in the library that provide access to data for researchers. A researcher is assigned to one of the computers, and the data is copied to the computer and placed in a designated folder. The computers are backed up once a week. Researchers can receive output results by placing the files in a specified folder on the computer, and request ICBS to send the data.

In both Tel Aviv and Haifa there are two computers in designated research rooms. The solution is much like Jerusalem, but the data is stored on removable discs encrypted with Microsoft bit locker. The removable discs are disconnected from the computer and stored in a safe place after work hours.

Bank of Israel - for 'research room' files and currently also for 'MUC' files

In the Bank of Israel there is a remote access research room with two computers. The room itself is protected by a smart card on the door and security cameras in the physical room. Researchers are identified by fingerprint, smartcards and a password before having access to servers and data located inside ICBS. The communication is protected by encryption in a site to site VPN connection on a leased line. Furthermore, firewalls, IDS/IPS and log monitoring systems are protecting the communication line.

Remote access facility – 'MUC' files

During the years 2008 to 2015, ICBS had a remote access service for researchers. Via this service, researchers were able to access and analyze data remotely. Output was supposed to be vetted by the ICBS for confidentiality, though in practice, this was not carried out.

The solution was implemented as a Virtual Desktop Infrastructure on technology from VMWare. The virtual environment ran on two physical servers with a shared storage system.

Researchers connected to the remote access solution through a secure communication server, Juniper, and were presented to a Windows XP desktop, which included the statistical programs SAS, SPSS, Excel and STATA. Each time a researcher logged off, the VDI machine was deleted. During the next logon, the researcher received a new VDI machine from scratch. Data was located on a file-server (data control server).

The remote access facility crashed in 2015 because of faulty discs in the storage system, and budget constraints which did not allow the purchase of new discs. This caused an unrecoverable error, and the service has been offline ever since. However, even before then, the system was unstable.

3. Conclusions and Recommendations

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

3.1. Organization, Workflow, Business Plan etc.

The meetings during the mission indicated that better coordination within the ICBS could benefit the progress within this component.

3.1.1 Establishment of a Working Group and a Steering Committee

The MS Experts recommend the establishing of a small inter-institutional working group that ensures progress in the component between missions. The working groups should include representatives of significant partners such as:

- International Relations and Statistical Coordination Department
- ICBS Legal Advisory Board
- Members of the Confidentiality Committee
- Representatives from Subject Units

It would be helpful if a Steering Committee was established in order to agree on the *Terms of Reference and a time-table for milestones* etc. related to the work of the working group. The MS experts suggest the following to be among the tasks for the working group:

- Organizational structure including organizational implementation plan
- Leaning of the present workflow and infrastructure
- Automation of the workflow
- Business model including price structure and monitoring of time used in each research project
- (Mapping of data available within ICBS and from other National Institutions)

- (The possibilities for general agreement for transmission of data from other Ministries to ICBS to be used for research)

The Steering Committee should preferably be established at a high level meeting of directors from the ICBS.

In order to ensure the continued progress, it is recommended to plan meetings on a regular basis between the parties involved, and it is further recommended that the working group inform the steering committee of their progress on a quarterly basis.

It is recommended to contact relevant researchers in order to receive their expectations relating to research services provided by ICBS.

3.1.2 One Dedicated Unit Responsible for Coordination of Research Services

In ICBS there are many internal players involved in each research project; however leadership and coordination seems to be missing. Therefore, it is recommended to make an organizational reorganization with one designated unit that will be responsible for coordination and tracking all processes involved in research projects.

3.1.3 Leaning of Workflows

The division of labor and the workflows concerning research service has to be agreed upon within the ICBS. Some confusion seems to be present. Far more transparency is needed regarding Research Service in ICBS.

Several work processes/workflows needs to be more simple and lean. It is suggested to consult with *Annex B1 – 14 and B1 - 15* as starting points. In addition, some work processes/workflows should be made automatic if possible.

3.1.4 Improving Documentation of the Time Used on Research Services

At present very little is known about how much time ICBS actually spends on providing services to researchers - and where in the organization the time is spend. The general view seems to be that the resources used on providing services to researchers far exceeds the costs that are covered by the researchers. However, improving documentation of the resources used on research services is needed. That is, resources used on specific research projects as well as resources used on general administration and knowledge-sharing concerning services to researchers. The experts recommend that a system for documenting the actual time used on research services all over the organisation is decided upon and implemented as soon as possible.

3.1.5 Drafting a Business Plan

In continuation of an improved documentation of the resources used on research services the experts recommend to make:

- Qualified calculations on research service costs as well as income from the research services in order to make ICBS able to decide on the yearly economic result of the overall Research Service activities
- A draft for a business plan

Management should be able to decide in advance whether they are willing to accept a given deficit or not. Management should decide on the size of the overhead percentage they want from research services. Overhead can be used for cost that is not directly associated with individual research projects such as e.g. development, certain administrative work etc. Furthermore, the future price policy should

be considered and decided upon. Thus, a more complete picture could be made of the economy for a future research service unit.

An important aspect of the link between cost and income related to research services in ICBS seems to be that resources are not linked directly to the actual work tasks. Several players describe that they take part in many work tasks related to research services but have no - or only few – resources to perform them. The experts strongly recommend that resources are allocated to the persons and/or subject divisions who actually carry out the work.

3.1.6 Metadata

It seems that much tailor made documentation is done today on data delivered to the specific research projects. In order to reduce this time consuming work, the experts recommend that the general process of improving metadata at ICBS is strengthened in years to come. This could significantly reduce the time used on tailor made documentation.

3.1.7 Legal Framework

It is recommended that a Legal Adviser actively participate in the working group for enhancing the research services in order to ensure that suggested changes are in accordance with relevant legislation and if not, initiate a process including a work plan for any legal changes needed.

3.2 Design IT Solution based on the Experience of Existing Remote Access

First of all, it may be a good idea to write down the primary goals for the design of a remote access solution. The goals can reflect parts as usability, security, capacity etc. It is further recommended to contact some of the primary researchers in order to gather information from the customers view. What does ICBS want to achieve by creating a remote access system?

During the mission there was a strong agreement on a "one size – fits all" design. The design must at least cover the different type of existing solutions with research rooms and remote access to MUC files. Even the research rooms can in the future be based on the remote access system.

The design should meet requirements for scalability in order to be able to adopt future demands for capacity and functionality. This may include:

- Enough bandwidth on communication lines
- Adopt growth in data
- Resources for application servers
- New functionality or statistical applications

A decision on the statistical software for the researchers should be made soon. Software like SAS is flexible but very expensive and can put pressure on the budget year after year. Free software such as R. should be considered. At Statistics Denmark researchers are asking for R and the number of R users is increasing. It might be a good idea to ask researchers what software they want to use, but this must be done carefully as it might result in more software than the budgets can afford.

Data can typically be presented either as files or in a database model. The existing solution for researchers in ICBS offers access to files. In the future, data could be stored in a database system instead of a fileserver. File based access is easy, while databases typically are more flexible.

The servers in the design will be using some kind of virtualization model. The solutions for Bank of Israel and the old remote access are based on virtualization of workstation computer on VMWare ESX, while Statistics Denmark is using Microsoft Remote Desktop Services (Terminal Server). A strategy with solutions like Remote Desktop Services might be more flexible and easier to scale.

Securing the traffic from researcher client computers and the servers/data in ICBS can be based on a security communication frontend solution, taking care of authentication of researcher's identities and encryption of data. In the old remote access system, the Juniper server is the frontend server for security. Because ICBS doesn't have control of the client computers, it is important to choose a solution with minimum requirement for software on the client.

After choosing the design, it is recommended to implement a small proof of concept/pilot in order to get an idea if the design can meet the specified requirements.

Development of Application for Automatically Controlling the Output Data

In the future ICBS will have more users than today, and researchers will request much more output than today. This might give a heavy resource load on the staff doing the manual controlling of the output. To reduce the extra cost it is recommended to design and develop a system for automatic control of output data, or at least part of the output data.

If output data meets specified requirements, an automated output control system can assist the manual checking. Otherwise output data will have to be controlled manually by the staff at ICBS, and the researcher must wait for this process to complete. The waiting time might be reduced by establishment of a output checking team rather than being depended on a single staff member.

IT Service Desk for Researchers

With more researchers having access to the remote access solutions, there may be more requests for technical support if there are problems when connection is made from a researcher client computer. It is recommended to have staff at ICBS supporting researcher and answering the special technical questions researchers might have. At Statistics Denmark the internal IT service desk has the skills to provide support to researchers in cases of technical problems.

It seems to be worthwhile considering implementing a Helpdesk service system to support the processes of providing services for researchers. The system could help the staff get an overview of incoming requests, and ensure that no request is forgotten or lost internally in the organization. This system can also be used in other organizational units at ICBS.

Table 1: *Recommended actions needed for moving forward as well as for preparing the next missions. For each action the experts recommend that ICBS agree on a realistic deadline and a responsible person/player for each action. The component BC leader should take action on this matter.*

Action	**Deadline**	**Responsible person/player**
A working group and a steering committee for research services established		
Draft for leaning of workflow produced		
Workflow steps to be automated identified		
A system for documenting the actual use of resources on research services decided upon		
A business plan drafted		
Plan for organizational reorganization presented to the steering committee		
Contact made with selected researchers to get input about		

their needs and wishes for a future remote access facility and services		
Server virtualization model selected (e.g. VMWare VDI, MS RDP)		
Decision made on the statistical software to be offered to the researchers		
Draft made of IT design of the remote access solution		
Draft list of requirements made for automated output control		
Considerations drafted on the need for a servicedesk for researchers		