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.2
Technological infrastructures and information security

Implemented by

- Mr. Povl Valeur Head of Division, IT, Statistics Denmark. pov@dst.dk
- Mr. Bo Guldager Clausen Chief Advisor Division of IT, Statistics Denmark, bgc@dst.dk

Jerusalem

June 6-9 2016

Version: Final

הלשכה המרכזית לסטטיסטיקה
Central Bureau of Statistics
دائرة الإحصاء المركزية

STATISTICS DENMARK

Authors' names, addresses, e-mails

Mr. Povl Valeur, Head of Division – IT,
Statistics Denmark
Sejrøgade 11
DK-2100 Copenhagen Ø
Denmark
Tel: +45 39 17 36 89
Email: pov@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):

Annex B2 - 1 Terms of Reference for B2
Annex B2 - 2 Meeting Program for B2
Annex B2 - 3 Persons Met at B2
Annex B2 - 4 Progress Since Mission B1 in April 2016 (BC Presentation)
Annex B2 - 5a ICBS Visions for Future Research Rooms and MUC Remote Access (BC Presentation)
Annex B2 - 5b ICBS Visions for Future Research Rooms and MUC Remote Access (BC Presentation)
Annex B2 - 6 IT Security Policy and Recommendations - ICBS (BC Presentation)
Annex B2 - 7 Security Commitments of Researchers and Research Institutions/Units - ICBS (BC Presentation)
Annex B2 - 9 IT and Data Security Measures for Implementing a RAS in Denmark (MS Presentation)
Annex B2 - 10 Example of an Automated System for Communication from DST (MS Presentation)
Annex B2 - 11 Sending Output and Output Control (MS Presentation)
Annex B2 - 12 Solution 1A Design Terminal Server (Figure - output of the activity)
Annex B2 - 13 Solution 1B Design VDI (Figure- output of the activity)
Annex B2 - 14 Solution 2 AtomNetwork RAS ICBS (Figure- output of the activity)

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 second mission of Component B – "Micro Data services to Researchers" 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 Technological Infrastructures and Information Security.

Based on a number of presentations and meetings with relevant staff members of the ICBS - the current situation on technological infrastructures and on research services at ICBS – as well as in Denmark and in other relevant countries - was described, discussed and assessed.

One of the main achievements of the activity was the creation of two drafts of the following potential technological infrastructures solutions for researcher's remote access to microdata from ICBS

- *Solution 1: A network design based on the solution used at Statistics Denmark*
- *Solution 2: A solution based AtomNetwork.*

Both solutions have some flexibility in how sub-components in each solution can be implemented such as e.g. flexibility in server virtualization model, software to be offered to the researcher etc. Pros and cons of each solution and its sub-components were discussed during the missions and summarized in the present report.

From these discussions the MS experts came up with the following recommendations:

- *To make a decision between continuing with a network design based on the solution used at Statistics Denmark for researchers remote access to microdata (Solution 1) or build a solution based AtomNetwork (Solution 2)*

- *To decide on the server virtualization model at an early stage (e.g. the software technology VMWare VDI or MS RDP) since this will have an impact on other aspects of the solution.*

- *To carefully select the statistical software to be offered to researchers based on input from the research community. Especially, since the decision about the available statistical software will have a big economic impact on ICBS and therefore should be taken with caution.*

- *To make a detailed draft of the IT design of the remote access solution and draft a list of requirements for automated processes.*

- *To carry out a thorough risk-analysis including all subcomponents as part of the process of designing a solution for remote access to microdata for researchers in close dialog with the the High Level Steering Committee on Data Security.*

- *To make a more accurate budget of costs for establishment and maintenance of the different components of the solutions outlined in this report from – both software and hardware, including the cost of installation and configuration.*

- *To draft a business plan for the selected solution when it comes to establishment as well as running costs for the first five years including a draft of who is going to pay for what e.g. external funding from the Ministry of Higher Education, the users (researchers), ICBS and other sources.*

- *To continue the great work on the possibilities to receive external financial support from the Ministry of Higher Education that will cover the costs of establishing and maintaining a new IT system for providing researchers with remote access to microdata.*

- *To create an IT service-desk for researchers.*

Finally the MS Experts would like to acknowledge the amount of work that the ICBS staff has put into some of the recommendations made in the first mission of this component.

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

The purposes of the mission were to:

- *Outline initial recommendations on strategic technological objectives.*

- *Prepare a draft for an automated system for communication between IT Servicedesk and research service.*

- *Prepare a detailed overview and time schedule (including deadlines) of the remaining activities in the IT component.*

The position analysis assisted the ICBS and the Twinning Project experts in getting a more detailed insight of the present situation regarding services to researchers and research environments in Israel, including IT-security, workflows, legal and technological aspects.

The MS experts made live demonstrations of the Danish solution for remote access to researchers, internal administration of projects and researchers and finally demonstrated how workflows have been implemented at Statistics Denmark.

Strategic discussions on IT-security and design of remote access in Israel were the key topics during the mission.

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 B2 - 1)*:

ICBS presented the current status of researcher access to microdata, IT security and workflows. The presentations included:

- *Progress since Mission B1 in April 2016 (Annex B2 - 4)*
- *Visions for Future Research Rooms and MUC Remote Access(Annex B2 – 5a,b)*
- *IT Security Policy and Recommendations Considered Necessary for Implementing a Remote Access System in Israel - for MUC and Research Room Files (Annex B2- 6)*
- *Security Commitments of Researchers and Research Institutions/units, which Need to be Signed Today and Plans for the Future (Annex B2 - 7)*

Solutions for researchers, IT security and implementation of workflows. The presentations included:

- *IT and Data Security Measures Considered Necessary for Implementing a Remote Access System in Denmark (Annex B2 - 9)*
- *Practical Live Demonstration of the Danish Research System.*
- *Presentation of Necessary Forms for Approval of Researchers by Statistics Denmark.*
- *Example of an Automated System for Communication from DST (Annex B2 -10)*
- *Sending Output and Output control (Annex B2- 11)*

The BC and MS staff had a study visit to the Ministry of Education to learn about their remote access system to educational data.

BC and MS discussed plans for a network design of the future system for remote access in ICBS. For output see Annex B2 – 12, 13, 14.

Based on these presentations, discussions on mission, vision and objectives with a focus on workflows and IT Solutions took place.

2.1 Assessment of the current IT development plan

BC presented a high-level vision (Annex B2 - 5a and 5b) which looked very well thought out and promising for the future.

In order to make the vision operational additional work needs to be done in order to detail the different parts of the vision.

2.2 Outlining initial recommendations on strategic technological objectives

2.2.1 Network design

During the mission the network design was discussed with the IT- and the IT security staff. It was obvious that IT security issues are treated differently in the ICBS and in Statistics Denmark.

In Denmark IT security issues are measured against business needs, and against the willingness of the top management is to take risks. There will often be solutions with a high level of security, where the remaining risks are described and well known. In the ICBS IT security issues often seem to block the business needs, but the MS consultants are aware of the differences in Israel and Denmark when it comes to security in general.

It is recommended to have all design proposals evaluated by the High Level Steering Committee on Data Security before making final decisions. The Committee consists of nine ICBS Staff members as well as members from the governmental agency that oversees cyber protection.

In the solutions, applications and data are physically located in the server room at ICBS in Jerusalem, and outputs can be sent by email after being controlled by staff members of the ICBS

2.2.2 Proposal for Network design based on the solution used at Statistics Denmark for researchers remote access to microdata - Solution 1

The first proposal for a design is a solution based on the current Danish design for remote access, but adjusted to conditions in the ICBS. On the network map, there are two versions of the solution, one in which is virtualization based on the terminal server (Annex B2 – 12), while the other uses VDI (Annex B2 – 13). The Danish system for remote access is based on a Microsoft terminal server used as a virtualization model (MS RDP), because it is easy to manage even with several hundred simultaneous users; however other solutions such as VMWare VDI are also a possibility.

Access to research room files from the research centres - presently Jerusalem, Tel Aviv and Haifa - and remote access to MUC files is based on the same solution. Hardware in research rooms will be thin clients acting as “terminals” connecting to research network physical placed at ICBS in Jerusalem.

2.2.2.1 Network segmentation – Solution 1

In the proposed solution the existing Internet network is protected by a firewall, where data can be copied to the research network directly from the internal network. No traffic should be allowed to be initiated from the research network to other networks including the internal network.

The DMZ network[1] is used, among others, to provide a secure access, including e.g. a token system or fingerprinting for multifactor authentication. This is the only network exposed to

[1] In computer security, a **DMZ** or **demilitarized zone** (sometimes referred to as a **perimeter network**) is a physical or logical subnetwork that contains and exposes an organization's external-facing services to a usually larger and untrusted network, usually the Internet. The purpose of a DMZ is to add an additional layer of security

the internet. Traffic from remote access users is terminated at the secure access solution, and forwarded to the research network. There are a number of different vendors of remote access solutions, and a key parameter in the selection process may be if there is local support in Israel. The previous system provided for researcher's remote access to MUC files until a few years back used a solution from a company named Juniper.

Management and maintenance etc. of the suggested solution for the research environment, including the research rooms, can be done from the existing external network by the IT Department in close dialog with a Research Service Unit.

In the suggested solution, networks for research rooms in Jerusalem, Tel Aviv and Haifa can connect to the research network using IPSEC VPN[2] connections over the internet. All research rooms are treated equally to standardize the design, though the research room in Jerusalem could be connected directly to the network infrastructure.

All servers for the research solution are connected to a single network, named research network. Here are servers for VMWare virtualization (or other virtualization Software models), Microsoft Terminal Servers, NAS (file server), databases and Active Directory.

2.2.2.2 Data security – Solution 1

In case the ICBS have sufficient specialized knowledge and staff resources for administrating database servers, implementation of a database solution to provide researchers with raw data could be further investigated in order to determine if such a solution will meet the security requirement of ICBS. A database solution will have the advantages of a lower demand for space and can provide new versions of data to all researchers in a single process, if all researchers access their data through the same tables.

In the suggested NAS[3] solution, each project should have a new set of files, even though much of the data is identical to the former version of data. Although raw data is placed in databases, researchers can still choose to use the file-based approach to work with data on the NAS system.

Both the database solution and the NAS solution for access for researchers are separated for each research project, and users logged-on to one project can't access data in another project at the same session. If a researcher is working in two or more projects, the researcher should have separate windows accounts with different multifactor authentications for each project. This will make it possible to block access between research projects.

To streamline administration and configuration of users, servers and computers on the research network, implementation of a dedicated active directory is recommended. If all devices and users on the research network are configured in an active directory, the daily monitoring, maintenance and management of the overall system will be more effective, and creation of automated administrative workflows will be easier.

---

to an organization's local area network (LAN); an external network node only has direct access to equipment in the DMZ, rather than any other part of the network.

[2] Internet Protocol Security Virtual Private Network

[3] Network-Attached Storage

2.2.3 AtomNetwork – an alternative solution for researchers remote access to microdata - Solution 2

An alternative and **second solution** discussed during the mission was a solution where the research network is separated into a number of smaller networks (Annex B2 – 14). There was a strong desire among the IT and security staff at the ICBS for such a solution since in their view this solution will improve the security level. The MS consultants agreed that the security in theory will be improved by making the split into multiple network segments, but as the solution becomes more complex, the risks of mistakes and configuration errors will be significantly higher.

In this second solution the firewall connecting all the networks, must be configured to allow necessary traffic so the solution can function with active directory, database access, file access etc.

Reflections on virtualization and data security are the same as described in the previous section of this report.

2.2.3.1 Network segmentation – Solution 2

In this second solution with segmentation of the research network the existing internal network is protected by a unidirectional switch only allowing traffic from the internal network to the research network using a very safe technique (Annex B2 – 14). There is an outgoing connection from the internal network to the research network, but only one physical switch port will be active at a time. It will not be possible to transfer files from the research network to the internal network.

DMZ network has the same functionality as the first design proposal.

The existing external network is used for emailing output control etc.

Active directory servers are isolated on a dedicated network, but must have access to most of the other networks, which are dependent on access to the domain controllers in the active directory.

Data is stored in three different networks, depending on the confidentiality of the data. One network for PUF files, one network for MUC files and a third data network for research room files.

Networks for research rooms have the same configuration as in the first proposal.

Management of components in the research networks, will be conducted from a management network, while there is a dedicated network for log consolidation.

2.2.4 Budgets for network designs

During the mission a draft budget was prepared based on the IT staff members' knowledge on the price level in Israel. It will be necessary to make a more accurate budget in order to take a

final decision. The budget contained only the prices of hardware and software, while the cost of installation and configuration are not included.

Selection of the solution used at Statistics Denmark or the AtomNetwork solution has only little impact on the budget for hardware and software, as it will be the same devices in both solutions. Implementation costs will rise in the AtomNetwork model since it is more complex.

Selection of the software that the ICBS will provide for researchers can increase the budget significantly. If SAS software is implemented in a server solution, this will increase the budget by up to 1,000,000 NIS, while initial license costs for R is 0 NIS. According to the Twinning contract, ICBS will work with the research community to better understand their needs and expectations of the system.

2.3 Producing a draft for an automated system for communication between IT service desk and research service

The Danish model for communication between IT-service desk and Research Service was discussed and it was emphasized again (as also done in Activity B1) that as much automation as possible should be done in order to keep a cost-effective system.

In Denmark a lot of systems have been built from scratch over the years by internal programmers, but in the long run this approach has proved to be costly to develop further. Statistics Denmark is currently looking into a common tool that supports workflows that can be built upon.

Based on experience and an internal evaluation at Statistics Denmark, Microsoft Sharepoint currently seems like the most common and promising workflow engine since it is possible to interface with different internal applications and workflows. In order to make it even easier to customize the workflows with less programming, it is possible to get add-ons for Sharepoint like Nintax (which is what Statistics Denmark is currently looking into).

There are other workflow engines on the market, but the recommendation is, that if one of these alternatives is chosen, it should only be used as the backend, and the ICBS should then implement the different workflows on top of the chosen workflow engine.

The actual or rather revised workflows still need to be sorted out. As pointed out in B1, there are a lot of processes and it is not clear what the future division of labor is going to be and there is also a need for leaning the processes. In other words it is recommended that the workflows are described with use-cases/user-stories and actual flow diagrams. Before the actual implementation, it is recommended that a "LEAN-approach" is used on the workflows in order to make them as effective as possible. These discussions will be initiated in the upcoming activity in September 2016.

2.4 Overview of IT related benchmarks and time schedule

- Strategic Plan elaborated for providing researchers with access to micro-data *(February 2017)*;
- Organizational and technological implementation plan, including data security described *(August 2017)*;

- Formal organizational structure proposed including IT *(August 2017)*;

The collection of input and expectations from researchers will take place over the course of November/December 2016 and a study visit including IT related issues is scheduled for December 2016. The last mission entirely dedicated to IT is scheduled to take place in February 2017.

3. Conclusions and Recommendations

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

The overall vision of the future remote access solution should be further elaborated and the different subparts should be worked out in more details.

A decision on either to continue with a Network design based on the solution used at Statistics Denmark for researchers remote access to microdata (Solution 1) or a solution based AtomNetwork (Solution 2) should be taken.

Concerning the future IT solution, the MS experts recommend that the decision on the server virtualization model (e.g. VMWare VDI, MS RDP) is taken at an early stage since this will have an impact on other aspects of the solution.

A decision on the statistical software which shall be offered to researchers should be taken, but with great care since this potentially has a long term economic impact. Included in this decision should also be a possible split of budget between ICBS and researchers.

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.

IT-security is obviously a big issue when granting access to data and a thorough risk-analysis should be carried out once a solution design is decided upon.

The High Level Steering Committee on Data Security should be consulted as necessary during the design phase.

It is recommended to continue the great work on the possibilities for financial support for the establishment and maintenance of a new IT system for providing researchers with remote access to micro-data.

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